Monitoring the Structural Changes of pre-edited mRNAs upon Editosome Binding - Evidence for the Evolutionary Origin of RNA-Editing

Dem Fachbereich Biologie der Technischen Universität Darmstadt

zur

Erlangung des akademischen Grades
eines *Doctor rerum naturalium*
genehmigte Dissertation von

Dipl. Biol.
Wolf-Matthias Leeder
aus Darmstadt, Hessen, Deutschland

1. Referent: Prof. Dr. H. Ulrich Göringer
2. Referent: Prof. Dr. Gerhard Thiel

Tag der Einreichung: 31. Mai 2016
Tag der mündlichen Prüfung: 8. Juli 2016

Darmstadt, den 30. Mai 2016
D 17

Die vorliegende Arbeit wurde am Fachbereich Biologie der Technischen Universität Darmstadt im Arbeitskreis Molekulare Genetik (Prof. Dr. H. Ulrich Göringer) angefertigt. Die Arbeit wurde durch die Deutsche Forschungsgesellschaft (DFG) innerhalb des Sonderforschungsbereichs 902: „Molecular Principles of RNA-based Regulation“ gefördert.

Teile der Arbeit gingen in folgende Veröffentlichungen ein:

Leeder WM, Voskuhl S, Göringer HU. 2016. The 2D-structure of the mitochondrial RPS12 pre-mRNA is not affected by volume exclusion conditions. [in preparation]

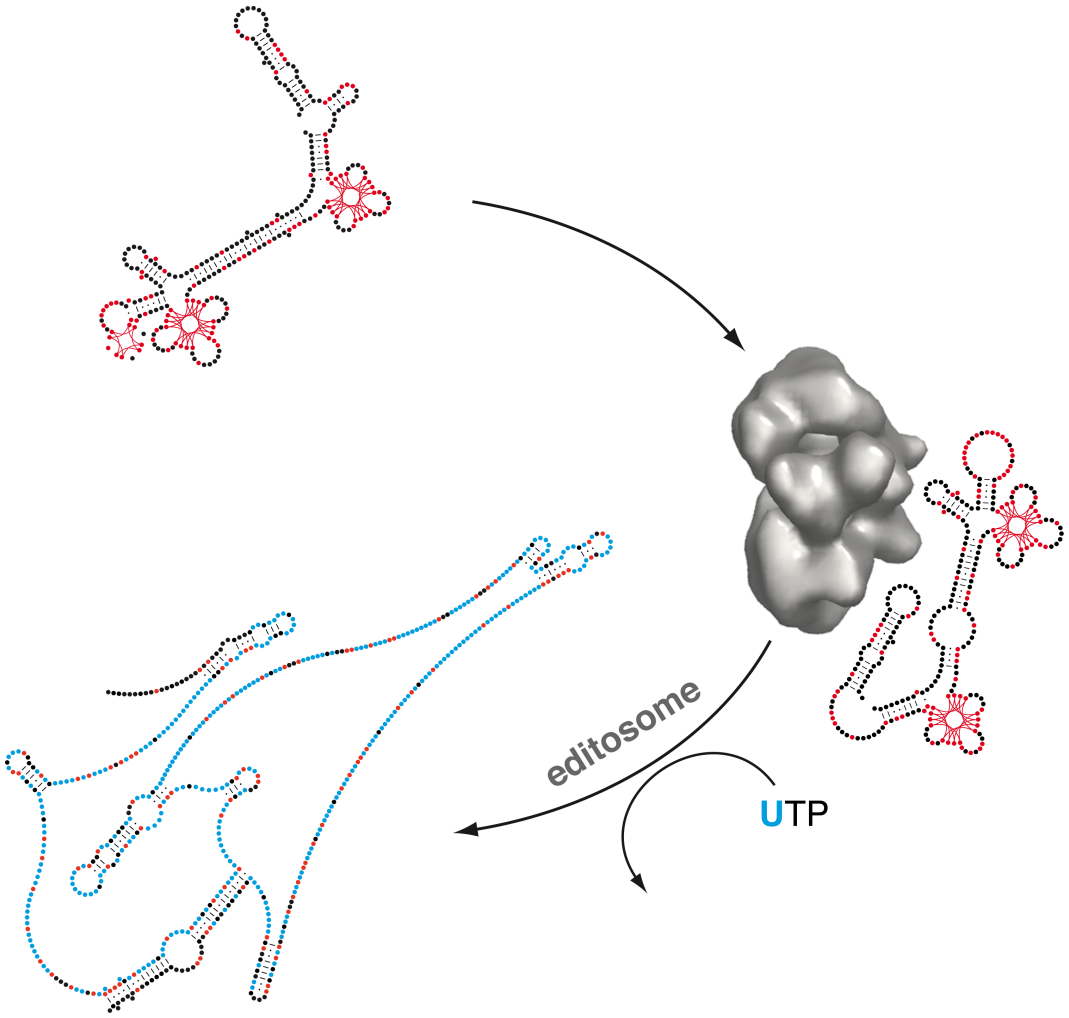
Leeder WM, Hummel NFC, Göringer HU. 2016. Multiple G-quartet structures in pre-edited mRNAs suggest origin of RNA editing in African trypanosomes. Sci Rep. 6:29810

Leeder WM, Voigt C, Brecht M, Göringer HU. 2016. The RNA chaperone activity of the *Trypanosoma brucei* editosome raises the dynamic of bound pre-mRNAs. Sci Rep. 6:19309.

Leeder WM, Reuss AJ, Brecht M, Kratz K, Wachtveitl J, Göringer HU. 2015. Charge reduction and thermodynamic stabilization of substrate RNAs inhibit RNA editing. PLoS One. 10:e0118940.

Contents

General introduction



The central role of RNA

In 1970 Francis Crick introduced the central dogma of molecular biology (Crick, 1970). Since then, RNA is predominantly known as a mediator in the gene-expression between the DNA-encoded genetic information and proteins. The 'one gene-one enzyme hypothesis' (Beadle & Tatum, 1941) and the 'one gene-one polypeptide hypothesis' (Ingram, 1957) were challenged in the 1980's culminating in the concept of introns and exons by Walter Gilbert (Berget *et al*., 1977; Chow *et al*., 1977; Gilbert, 1978). In a foresight, he ventilated the idea of 'introns being both frozen remnants of history and sites of future evolution', not ruling out other roles. In the course of the elucidation of the splicing-mechanism Cech and colleagues discovered the first auto-catalytic RNA (Kruger *et al*., 1982). Only one year after the discovery of self-splicing introns, Sidney Altman's group presented the first true RNA enzyme or ribozyme capable of multiple-turnover (Guerrier-Takada *et al*., 1983). RNaseP facilitates the maturation of transfer (t)RNA precursors by cleaving off the leader sequence from the otherwise non-functional 5'-ends ceasing the exclusive role of proteins as catalysts. Although these pioneering works underpinned the RNA world hypothesis (Gilbert, 1986; Müller, 2006), highly structured and/or catalytically active RNAs, such as tRNAs, Group I and Group II intros have been considered exotics (Novikova *et al*., 2013), while non-protein coding DNA was deemed as 'junk' that prokaryotes lost in the process of streamlining their genomes (discussed in Gilbert, 1985).

Since then, myriads of catalytic, structural and regulatory RNAs or RNA-motifs have been discovered, superseding proteins as the sovereigns of gene expression, development and genome maintenance (Sharp, 2009; Novikova *et al*., 2013; Dey *et al*., 2014; Cech & Steitz, 2014; Nam *et al*., 2016). Introns display a clear-cut function as a source for small nucleolar (sno)RNAs that facilitate ribosomal (r)RNA maturation (Lui & Maxwell, 1990; Osman *et al*., 2016). Alternative splicing expands protein-diversity and intron-derived microRNAs regulate gene expression (Ambros *et al*., 2003; Black, 2003; Osman *et al*., 2016). MicroRNAs guide an exonuclease-active protein complex to a specific region in an mRNA by Watson-Crick base-pairing that ultimately leads to the cleavage and degradation of the mRNA or translational repression (Bartel, 2004). A truly outstanding example for RNAs regulating RNAs in *trans* are circular-RNAs (Salzman *et al*., 2012; Lu *et al*., 2015). Circular-RNAs are characterized by 5'−3' linkage that makes them to resistant to exonuclease digestion. They are proposed to be a long-living 'sponges' that absorb microRNAs (Guo *et al*., 2014a). Next to functions arising from 1D-structure *i.e.* primary sequence, RNAs fold into sophisticated 2D and/or 3D-structures. Their versatile functions are inherently linked to these structures. In 2009 Weeks and colleagues demonstrated that the HIV-genome harbors a second layer of genetic information encoded in the 2D-structure of the genomic RNA. They showed that genes are organized in RNA-domains punctuated by extensive 2D-structure. They suggested that these RNA 2D-structures modulate translation and promote native protein folding (Watts *et al*., 2009).

The regulation of RNA by RNA in a ligand dependent manner made Jacob and Monod's hypothesis of proteins as sole biopolymers capable of acting as gene-repressor obsolete (Jacob & Monod, 1961). Riboswitches are a prime example for RNAs regulating in *cis*. They are capable to switch between two structures upon complexing small-molecule metabolites specifically. The structural transition either forms a terminator stem-loop or sequesters the start codon and/or the Shine-Dalgarno sequence to inhibit translation. The

sensing capabilities of the complex and flexible 3D-RNA-structures are not restricted to small molecules. T-Box riboswitches detect uncharged tRNAs to initiate a response to amino-acid deficiency (Grundy *et al*., 1994; Klein & Ferré-D'Amaré, 2006; Breaker, 2011; Mellin & Cossart, 2015; Zhang & Ferré-D'Amaré, 2015). Next to that, RNA thermometers measure temperature thereby controlling expression of virulence genes (Johansson *et al*., 2002). The diverse sensing and switching capability of riboswitches therefore empowers them as vital building blocks for the design of artificial genetic circuits (Sudarsan *et al*., 2006; Win & Smolke, 2008; Miyamoto *et al*., 2013). Aside from that, RNA can sense ligands as small as metal ions with high specificity (Paige *et al*., 2011; DasGupta *et al*., 2015). G-quadruplexes (GQ's) are four-stranded structures of hoogsteen-hydrogen bonded guanines chelating central potassium ions with the generic motif $G_{>2}N_{y1}G_{>2}N_{y2}G_{>2}N_{y3}G_{>2}$. In nature, GQ's carry out a wide variety of functions (Rhodes & Lipps, 2015). To name just a few, their functions range from ribosome-recruitment over splicing-enhancement to replication/transcription switches (Agaronyan *et al*., 2015; Bhattacharyya *et al*., 2015; Ribeiro *et al*., 2015). Due to their widespread and diverse functions in gene regulation, GQ's emerged as medically relevant targets (Wang *et al*., 2015). Their propensity to form macromolecular assemblies moved GQ's in the focus of supramolecular chemistry and nanotechnology (Biyani & Nishigaki, 2005; Davis & Spada, 2007; Guo *et al*., 2014b; Wang *et al*., 2014). DNA GQ's serve as 'caps' of eukaryotic genomes protecting the chromosomes from deterioration, which is of special interest in aging and cancer research (Shay, 2016; Tan & Lan, 2016). The machinery maintaining these structures is a ribonucleoprotein particle utilizing a flexible RNA-scaffold and a RNA-template for reverse transcription of repetitive DNA-GQ-elements (Zappulla & Cech, 2004). The snoRNA-mediated maturation of rRNAs expands the biochemical repertoire by the so called RNA-editing (McMahon *et al*., 2015). RNA-editing, which is in many cases the site-specific alteration of nucleotides or nucleobases (Göringer, 2008), frees RNA from the restricted four letter code climaxing in the most advanced RNA-catalyst known, the ribosome. The biological centrality of rRNA is underpinned by the fact that microorganisms have forged a broad arsenal of specific "bioweapons", such as aminoglycosides, targeting the rRNA instead of the ribosomal proteins that discriminates between friend and foe (Blaha *et al*., 2012). Next to that, bacteria have evolved an adaptive immune system. The clusters of regularly interspersed short palindromic repeats (CRISPR) system protects bacteria from invading phages and plasmids (Barrangou *et al*., 2007). It utilizes copies of genetic material acquired from previous invaders stored in the genome and inherited to the next generation. The RNA transcribed from these copies serve as a 'guideRNA' to target foreign DNA for destruction by the Cas9 nuclease. Almost instantly, the CRISPR system emerged as a novel and versatile tool for genome editing in molecular biology (Lander, 2016).

The most prominent example of a long non-coding (lnc)RNA is the first one discovered: Xist (Brockdorff *et al*., 1992; Brown *et al*., 1992). Xist stands for X-inactive-specific transcript. The RNA coats the X-chromosome from which it is expressed thereby triggering gene silencing that provides gene-dosage compensation between male and female mammals (Pandya-Jones & Plath, 2016). Since then, the number of lncRNAs is estimated to be more than 15000 in humans exalting the biological relevance of these RNAs (Derrien *et al*., 2012). In fact, over 75% of the human genome is

actively transcribed (Djebali *et al*., 2012). Non-coding RNAs account for ~80% of the transcriptome in humans (Kapranov *et al*., 2007). Taking into account that <2% of the genome is encoding proteins, the consideration of non-protein-coding DNA/RNA as junk or transcriptional noise is obsolete (Venter *et al*., 2001; De Lay & Garsin, 2016; Osman *et al*., 2016). The same holds true for RNA being a passive intermediate in bridging the genomically encoded 1D-information to the three-dimensional world. Recent studies have solidified the role of RNAs as the primary regulators of gene expression. The structure-function interplay of RNAs therefore gave rise to the large scale analysis of the 'RNA structurome' to elucidate the multifarious functions of the diverse RNAs. Recent models and reliable approximations of lncRNA-interactions suggest a mind-boggling scope of functions (Novikova *et al*., 2013; Cech & Steitz, 2014; Strobel *et al*., 2016). Even if life has not originated from an RNA world, higher eukaryotic life can be considered as such.

African trypanosomes

African trypanosomes are single cellular parasites and the causative agent of sleeping sickness. They are therefore of medical importance (Stuart *et al*., 2008). Trypanosomes belong to the order of kinetoplastida, which are early branching eukaryotes (Simpson *et al*., 2002). Kinetoplastids are characterized by an unusual arrangement of their mitochondrial genome that is organized in a supramolecular, disc-shaped assembly, the kinetoplast (Lukeš *et al*., 2002; Liu *et al*., 2005; Jensen & Englund, 2012). It consists of two classes of circular DNA molecules, the homogenous 23kb-large maxicircles and the ~1kbp sized, sequence-diverse minicircles. The single mitochondrion of African trypanosomes contains approximately 50 maxicircles and several thousand interlocked minicircles that are catenated into a massive DNA network (Laurent *et al*., 1971; Brack *et al*., 1972; Renger & Wolstenholme, 1972; Steinert & Van Assel, 1975; Kleisen *et al*., 1976a; Kleisen *et al*., 1976b; Englund, 1978, Jensen & Englund, 2012). Due to its complexity, the kinetoplast requires a sophisticated replication procedure preceding the nuclear DNA replication (Hoeijmakers & Weijers, 1980; Gluenz *et al*., 2011; Jensen & Englund, 2012). Unlike other mitochondrial genomes, over half of the genes lack viable sequence information. The primary transcripts therefore require a bizarre posttranscriptional maturation process, U-insertion/deletion RNA-editing (Benne *et al*., 1986; Feagin *et al*., 1988). Mitochondrial RNA-editing in kinetoplastids is characterized by the sequence specific insertion and/or deletion of exclusively U-nucleotides thereby converting cryptic pre-edited mRNAs into translatable transcripts (reviewed in Aphasizhev & Aphasizheva, 2011; Aphasizhev & Aphasizheva, 2014). In the most extreme cases hundreds of U-nucleotides are site-specifically inserted while several dozen U's are deleted from the primary transcripts. This accounts for up to ~50% of the coding information as well as the creation of start and stop codons. The editing reaction is catalyzed by an 800kDa poly-protein complex that, in analogy to the ribosome and spliceosome, has been termed the editosome (Pollard *et al*., 1992; Göringer *et al*., 1994; Köller *et al*., 1994; Corell *et al*., 1996; Peris *et al*., 1997; Rusché *et al*., 1997). The editosome catalyzes a multistep enzyme cascade (Blum *et al*., 1990; Sabatini & Hajduk, 1995; Frech & Simpson, 1996) that requires small non-coding trans-acting guide (g)RNAs (Blum *et al*., 1990; Blum & Simpson, 1990; Schmid *et al*., 1995; Hermann *et al*., 1997) acting as quasi templates for the RNA-editing reaction. Guide RNAs are almost exclusively encoded on the minicircles (Pollard *et al*., 1990; Sturm & Simpson,

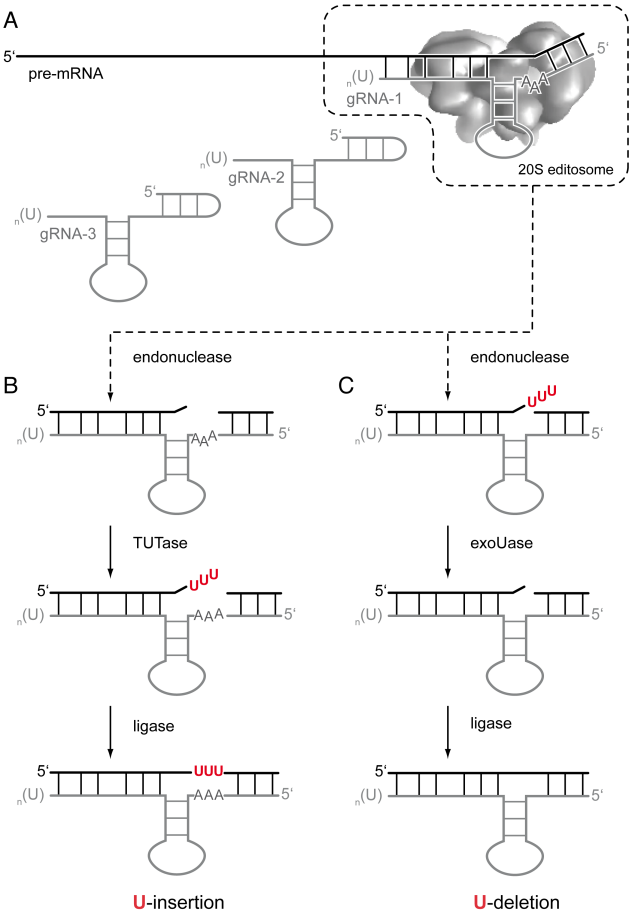


Figure 1: RNA-editing reaction-cycle. (Kruse *et al.*, 2013) A: The pre-mRNA (black) base-pairs with its cognate guide (g)RNA (grey) to form a three-way-helix junction. Guide RNAs are poly-uridylated and provide sequence information for multiple U-insertion and/or deletion editing-sites. In almost all cases several gRNAs (gRNAs 1-3) are required for the editing of one pre-mRNA. The 20S editosome (grey, dumbbell-shaped structure) provides the catalytically active surface for the reaction. B: Mechanistic outline of the U-insertion editing reaction. C: Basal reaction steps of the U-deletion RNA-editing cycle. Both reactions initiate with the endonucleolytic cleavage of the pre-mRNA at the first unpaired nucleotide upstream of the "anchor-region". An editing-site is defined by the "guiding" nucleotides of the gRNA. The U-insertion reaction proceeds with the addition of U-nucleotides from UTP to the 3'end of the upstream pre-mRNA fragment. The terminal-uridylyl transferase (TUTase) adds U's until base-complementarity with the guiding A's is achieved. The U-deletion reaction proceeds with an exonucleotlytic trimming reaction. Excess U's are removed from the 3'-end of the upstream pre-mRNA fragment by a 3'-5' exonuclease (ExoUase). The reaction is completed by the ligation of the two pre-mRNA fragments. Faithful editing of all editing-sites encoded by a certain gRNA creates the anchor-sequence for the next gRNA. The reaction therefore has a 3'-5' directionality.

1990; Pollard & Hajduk, 1991), while the mitochondrial genes are located on the maxicircle. The maxicircle harbors the genes of two ribosomal RNAs, a ribosomal protein, several subunits of the oxidative phosphorylation chain and genes of unknown function (Eperon *et al.*, 1983; Benne *et al.*, 1983; Hensgens *et al.*, 1984; Payne *et al.*, 1985; Benne *et al.*, 1986; Feagin *et al.*, 1987; Feagin *et al.*, 1988; Shaw *et al.*, 1988; Bhat *et al.*, 1990; Koslowsky *et al.*, 1990; Maslov *et al.*, 1992; Sloof *et al.*, 1992; Souza *et al.*, 1992). Some genes are developmentally regulated (Feagin *et al.*, 1987; Souza *et al.*, 1992; Corell *et al.*, 1994).

Trypanosoma brucei exhibits a life cycle involving a mammalian host and an invertebrate vector, the Tsetse fly (Stuart *et al.*, 2008). As a blood-borne parasite, the organism generates energy from glycolysis while it relies on oxidative phosphorylation in the insect stage. Therefore, RNA-editing is vital in the fly stage of the parasite. Though, both the U-insertion/deletion RNA-editing machinery and the kinetoplast are absent in the host, these unique characteristics of trypanosomes pose promising drug targets (Roy Chowdhury *et al.*, 2010; Salavati *et al.*, 2011; Jensen & Englund, 2012). Unfortunately, the long-standing search for anti-protozoal compounds has spawned only a few chemotherapeutics, which suffer from severe side effects (Steverding, 2010). The quest for more effective and less toxic trypanocidal drugs is hindered by the fact that many of the molecular details of the RNA-editing reaction and the mitochondrial genome organization are only partially understood. While the global structure and the individual components of the editosome are known, no conclusive data exists on the 3-dimensonal arrangement of these (reviewed in Göringer, 2012). Next to that, a chaperone like unwinding function of the editosome has been demonstrated but the underlying mechanism and biological function remained elusive (Böhm *et al.*, 2012). The situation is even more enigmatic on the

Editosome	Motif	Identified/Proposed function
MP18	OB fold	gRNA binding
MP19	OB fold	Interaction
MP24	OB fold	gRNA binding
MP41	U1-like	Interaction
MP42	Zn-finger, OB fold	Endo/exonuclease; organization
MP44	RNaseIII, Pumilio, U1-like	Editosome integrity
MP46	RNaseIII, Pumilio, U1-like	Editosome integrity
MP47	U1-like	Interaction
MP48/REL2	Ligase, tau, K	RNA ligase
MP49	U1-like	Interaction
MP52/REL1	Ligase, tau, K	RNA ligase
MP57/RET2	NZ, PAP-core, PAP-assoc	TUTase
MP61/REN2	RNaseIII, dsRBM, U1-like	Interaction-specific endonuclease
mHel61p	DExH/D-box Helicase	RNA helicase, RNPase
MP63	Zn-finger, OB-fold	Interaction
MP67/REN3	RNaseIII, dsRBM, U1-like	COXII-specific endonuclease
MP81	Zn-finger, OB-fold	Interaction
MP90/REN1	RNaseIII, dsRBM, U1-like	Deletion-specific endonuclease
MP99/REX2	5'/3' exonuclease, EEP-domain	Nuclease/nucleotidyl phosphatase
MP100/REX1	5'/3' exonuclease, EEP-domain	Nuclease/nucleotidyl phosphatase
Accessory Proteins		
gBP21	R-rich	RNA matchmaking
gBP25	R-rich	RNA matchmaking
RBP16	Cold shock domain, RGG	Interaction
TbRGG1	RGG	mRNA stabilization
KRET1	NT, PAP-core, PAP-assoc, Zn-finger	TUTase

Table 1. List of the editosomal protein inventory. (Göringer *et al.*, 2011) Most integral and accessory proteins are annotated as: MPxx: mitochondrial protein, molecular weight in kDa. Tb: *Trypanosoma brucei*. Sequence motifs: EEP: endo-exo-phosphatase; RNaseIII: endoribonuclease motif from RNaseIII; dsRBM: double-stranded RNA binding motif; U1-like: U1-like Zn-finger motif; Pumilio: Pumilio domain RNA binding motifs; ligase: signature ligase motif; tau and K: putative microtubule associated tau and kinase light chain domains; RGG: arginine-glycine-glycine motif; R-rich: arginine-rich domain; DExH/D-box: aspartate-glutamate-x-histidine/ aspartate helicase consensus sequence. The identified/proposed function is derived from experiments or sequence analysis. "Interaction": binding of RNA or protein, no catalytic activity has been described.

RNA level. Although a 3D-model of the gRNAs (Hermann *et al.*, 1997) and a 2D-structure of the gRNA/pre-mRNA hybrid exists (Reifur & Koslowsky, 2008), there is no knowledge concerning the structure of full-length pre-mRNAs not to mention the potential structural rearrangements upon mRNA/editosome interaction. The same holds true for the kinetoplast-replication. Although the process itself and several contributors have been described, fundamental questions remain unanswered. This not only includes how the compact structure of the kinetoplast is maintained or how mini- and maxicircles are equally distributed to the sister networks (Jensen & Englund, 2012).

U-insertion and deletion RNA-editing – an enzyme cascade

The U-insertion/deletion RNA-editing reaction is catalyzed by an 800kDa large multiple protein complex, which has been termed the editosome (Pollard *et al.*, 1992; Göringer *et al.*, 1994; Köller *et al.*, 1994; Corell *et al.*, 1996; Rusché *et al.*, 1997; Peris *et al.*, 1997; Golas *et al.*, 2009; Göringer *et al.*, 2011; Göringer, 2012). The reaction initiates with the recruitment of a gRNA and the formation of a gRNA/pre-mRNA hybrid that adapts a three-way-helix topology (Leung & Koslowsky, 2001; Yu & Koslowsky, 2006; Reifur & Koslowsky, 2008). The structure is stabilized by a short "anchoring" duplex directly downstream to the editing-site. At the editing-site, the gRNAs act as quasi templates dictating either the insertion or deletion of U-nucleotides (Fig. 1). Unpaired A's in the gRNA bordering the anchor region specify U-insertion events with UTP as a substrate, while non-base-paired uridylates in the pre-mRNA become deleted (Seiwert & Stuart, 1994; Frech & Simpson, 1996; Kable *et al.*, 1996; Seiwert *et al.*, 1996). The first enzymatic step is the endonucleolytic cleavage of the pre-mRNA at the first unpaired nucleotide upstream of the 'anchoring' duplex. Three related enzymes facilitate the endonucleolytic cleavage. All of them contain U1-like Zn-finger motifs, a RNaseIII domain and a dsRBM sequence. While TbMP90 is responsible for deletion-sites, TbMP63 is specific for insertion-sites (Carnes *et al.*, 2005; Trotter *et al.*, 2005). The function of the third enzyme TbMP67 remains not fully understood. At insertion sites, U's

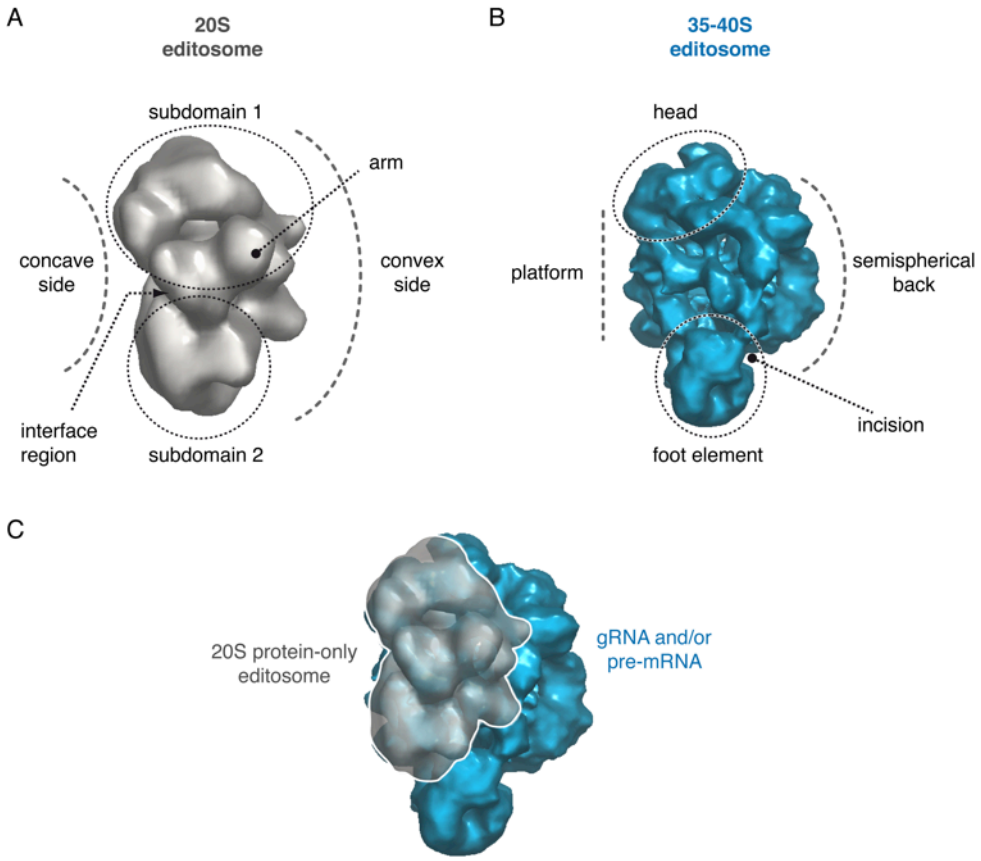


Figure 2: Structural landmarks of native *T. brucei* editosomes. (Golas *et al.*, 2009) A: Consensus 3D-structure of the protein-only 20S editosome. The two non-identical subdomains are linked by an interface region (EMDB-1595). B: Consensus 3D-structure of the RNA-loaded 35-40S editosome (EMDB-1594). The RNA-containing semispherical back and additional structural features are indicated. C: 3D-docking/alignment of the 20S editosome into the 35-40S particle. The protein-only 20S editosome can be integrated into the platform and head domain. The 20S editosome has no correlates in the foot and the semispherical back. The 30-40S particles most likely represent editosomes engaged in the process of RNA-editing.

are added to the 3'-end of the mRNA fragment by a poly(A)-polymerase-like enzyme executing terminal-uridylytransferase (TUTase) activity until base-complementarity with guiding A's is achieved (Ernst *et al.*, 2003). At deletion sites, the endonucleolytic cleavage is followed by a U-specific exonuclease (ExoUase) trimming reaction. The ExoUase removes excess U's from the 3'-end of the upstream mRNA fragment in a gRNA-depended fashion. ExoUase activity has been identified for two related proteins: TbMP99 and TbMP100 (Kang *et al.*, 2005; Rogers *et al.*, 2007). Both enzymes harbor *N*-terminal 5'-3' exoribonuclease and *C*-terminal exo/endo/phosphatase motifs. TbMP42 is a structure sensitive endo-exo ribonuclease. It has two C_2H_2-Zn-finger motifs, a potential oligonucleotide/oligosaccharide binding (OB) fold domain and likely follows a metal ion catalysis mechanism (Brecht *et al.*, 2005; Niemann *et al.*, 2008; Niemann *et al.*, 2009). All three nucleases perform nucleotidyl phosphatase activity (Niemann *et al.*, 2009). The enzyme cascade is completed with the ligation of the processed mRNA fragments that is executed by two editing-specific ligases, TbMP48 and TbMP52 (McManus *et al.*, 2001; Rusché *et al.*, 2001; Gao & Simpson, 2003; Deng *et al.*, 2004). Most gRNAs facilitate editing of several editing sites (Maslov & Simpson, 1992; Koslowsky *et al.*, 2014). Depending on the length of a given transcript, the complete editing of an extensively or 'pan-edited' pre-mRNAs can require >30 gRNA classes (Maslov *et al.*, 1994; Koslowsky *et al.*, 2014).

Structure, composition and function of the *T. brucei* editosome

Next to the core enzymatic activities described in the previous section, the editosome contains several accessory factors (Table 1). During the process of RNA-editing the sequence of the pre- and partially edited mRNA changes constantly. Next to that, the anchor duplex incrementally extends with every edited and therefore base-complementary editing-site. This necessitates permanent rearrangements of the pre-mRNA-gRNA hybrids three-way-junction. After faithful editing the base complementary gRNA has to be displaced in order to facilitate annealing of the next gRNA. Therefore, accessory factors are required such as matchmaking-type RNA/RNA annealing factors gBP21 and gBP25 (Köller *et al*., 1997; Allen *et al*., 1998; Müller *et al*., 2001; Blom *et al*., 2001; Müller and Göringer, 2002; Aphasizhev *et al*., 2003a; Schumacher *et al*., 2006), the proteins RBP16 (Hayman & Read, 1999; Pelletier *et al*., 2000; Miller & Read, 2003; Pelletier & Read, 2003), TbRGG1 (Vanhamme *et al*., 1998; Hashimi *et al*., 2008), REAP1 (Madison-Antenucci *et al*., 1998; Madison-Antenucci & Hajduk, 2001; Hans *et al*., 2007), and RBP38 (Sbicego *et al*., 2003). Next to that, the 3'-end-specific TUTase KRET1 (Aphasizhev *et al*., 2002; Aphasizhev *et al*., 2003b; Aphasizheva & Aphasizhev, 2010) and the mitochondrial DExH/D protein mHel61p (Missel *et al*., 1997; Kruse *et al*., 2013) have been described as accessory factors. The global structure of the editosome has been elucidated from steady-state preparations by transmission electron microscopy (TEM) and cryo-EM revealing two distinct classes of high-molecular-mass assemblies (Golas *et al*., 2009). The two classes consist of large, irregular shaped complexes. The smaller class has dimensions of 21 x 26nm while the larger one is over 26nm in diameter. Both assemblies are characterized by distinct 3D-structures (Fig. 2). The two classes represent the protein-only and the RNA-complexed editosomes that show sedimentation-coefficients of 20 and 35-40 Svedberg-values (S) (Pollard *et al*., 1992; Corell *et al*., 1996; Golas *et al*., 2009). While the 20S editosomes are competent of faithfully editing synthetic substrates, the 30-40S particles are not (Igo *et al*., 2000; Igo *et al*., 2002; Carnes & Stuart, 2007; Golas *et al*., 2009). Most likely this results from the RNA-binding site being blocked by endogenous gRNAs and/or pre-mRNAs. 3D-docking and alignment approaches suggest that the head and platform domain of the 35-40S particles represent the protein-only 20S editosomes. Thus the semi-spherical back of the 35-40S consists of bound gRNAs and/or pre-mRNAs, which is the biochemical difference between the two classes. This scenario has been confirmed by interconversion experiments. RNase-digestion converts 35-40S editosomes to 20S editosomes, while incubation of isolated 20S editosomes with *T. brucei* mitochondrial RNAs generates 35-40S editosomes (Golas *et al*., 2009).

The 20S editosome binds RNA with high affinity (Golas *et al*., 2009; Böhm *et al*., 2012; Katari *et al*., 2013). The equilibrium dissociation constants (K_d) are in the low nanomolar concentration range. Experimental data indicates that the 20S editosome has a bi-functional reaction center with one RNA-binding site (Böhm *et al*., 2012). The hypothesis is supported by the fact that the core enzymes are present at least as pairs. Furthermore, the 20S editosome displays two similar but not identical subdomains linked by an interface region (Golas *et al*., 2009). Next to that the synthetic deletion- or insertion RNA-editing substrates act as competitive inhibitor in a reciprocal fashion. In addition, TEM images of 20S editosomes interacting with gold-labeled RNAs confirms a single RNA-binding site. Interestingly, the 20S editosome executes an ATP-independent RNA-unwinding reaction, which was

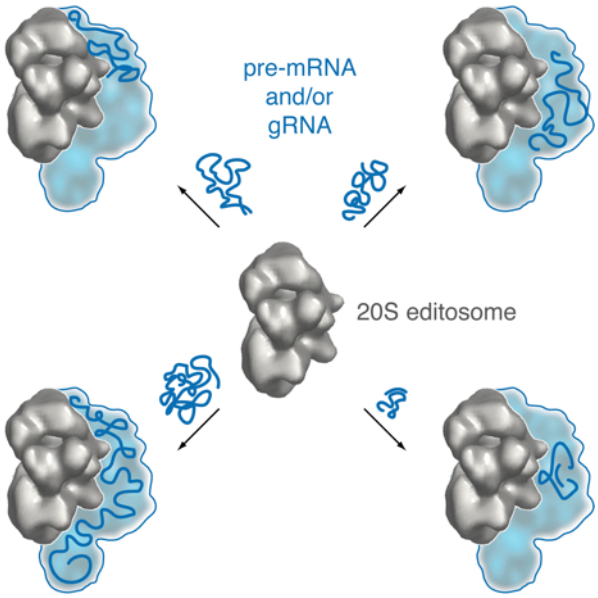


Figure 3. 20S editosome to 35-40S conversion. The protein-only 20S editosome (grey) binds pre-mRNA and/or gRNAs (blue lines) thereby forming the 35-40S editosome. The consensus structure of the RNA-component of the 35-40S particle is depicted as a blue, semitransparent envelope. The AFM-detected RNA-unfolding capability of 20S editosomes and the 20S to 35-40S interconversion experiments suggest that the semispherical back represents an ensemble of partially unfolded RNAs.

uncovered by incubation of steady-state 20S editosomes with a panel of mitochondrial transcripts. Atomic force microscopy (AFM) revealed multiple editosomes bound to unfolded mRNAs. The amount of unfolded mRNAs and mRNA-editosome assemblies increased in a time dependent fashion (Böhm *et al.*, 2012). This suggests that the semispherical back is partially unfolded RNA, which interacts with large surface patches of the 20S editosome (Fig. 3). Even though the protein inventory and the global structure of the editosome have been solved, almost no efforts have been made to elucidate the structure of the mRNA. However, AFM imaging shows mitochondrial transcripts as monodisperse populations of small, roundish structures (Böhm *et al.*, 2012). For example, the ~400nt long pre-edited mRNA coding for ATPase subunit 6 (A6) appears as a half-ellipsoid with an average diameter of 13nm and a height of ≤2nm that suggests a highly folded RNA.

The RNA 2D- and 3D-structure and its remodeling upon protein interaction and/or ribonucleoprotein assembly plays a key role in gene regulation and expression (Weeks & Cech, 1996; Zappulla & Cech, 2004; Abeyrathne *et al.*, 2016; Fischer *et al.*, 2015; Hardin *et al.*, 2015; Sun *et al.*, 2015; Wahl & Lührmann, 2015; Agafonov *et al.*, 2016). Recent progress in chemical footprinting methods allows RNA 2D-structure elucidation *in vitro* and *in vivo* (Duncan & Weeks, 2010a; Duncan & Weeks, 2010b; McGinnis & Weeks, 2014; Smola *et al.*, 2015).

References

Abeyrathne PD, Koh CS, Grant T, Grigorieff N, Korostelev AA. 2016. Ensemble cryo-EM uncovers inchworm-like translocation of a viral IRES through the ribosome. Elife. 5. pii: e14874.

Agafonov DE, Kastner B, Dybkov O, Hofele RV, Liu WT, Urlaub H, Lührmann R, Stark H. 2016. Molecular architecture of the human U4/U6.U5 tri-snRNP. Science. 351:1416-1420.

Agaronyan K, Morozov YI, Anikin M, Temiakov D. 2015. Mitochondrial biology. Replication-transcription switch in human mitochondria. Science. 347:548-551.

Allen TE, Heidmann S, Reed R, Myler PJ, Göringer HU, Stuart KD. 1998. Association of guide RNA binding protein gBP21 with active RNA editing complexes in *Trypanosoma brucei*. Mol Cell Biol. 18:6014-6022.

Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D. 2003. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. Curr Biol. 13:807-818.

Aphasizhev R, Sbicego S, Peris M, Jang SH, Aphasizheva I, Simpson AM, Rivlin A, Simpson L. 2002. Trypanosme mitochondrial 3' terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing. Cell. 108:637-648.

Aphasizhev R, Aphasizheva I, Nelson RE, Simpson L. 2003a. A 100-kD complex of two RNA-binding proteins from mitochondria of *Leishmania tarentolae* catalyzes RNA annealing and interacts with several RNA editing components. RNA. 9:62-76.

Aphasizhev R, Aphasizheva I, Simpson L. 2003b. A tale of two TUTases. Proc Natl Acad Sci U S A. 100:10617-10622.

Aphasizheva I, Aphasizhev R. 2010. RET1-catalyzed uridylylation shapes the mitochondrial transcriptome in *Trypanosoma brucei*. Mol Cell Biol. 30:1555-1567.

Aphasizhev R, Aphasizheva I. 2011. Uridine insertion/deletion editing in trypanosomes: a playground for RNA-guided information transfer. Wiley Interdiscip Rev RNA. 2:669-685

Aphasizhev R, Aphasizheva I. 2014. Mitochondrial RNA editing in trypanosomes: small RNAs in control. Biochimie. 100:125-131

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 315:1709-1712.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 116:281-297.

Beadle GW, Tatum EL. 1941. Genetic Control of Biochemical Reactions in *Neurospora*. Proc Natl Acad Sci U S A. 27:499-506.

Benne R, De Vries BF, Van den Burg J, Klaver B. 1983. The nucleotide sequence of a segment of *Trypanosoma brucei* mitochondrial maxi-circle DNA that contains the gene for apocytochrome b and some unusual unassigned reading frames. Nucleic Acids Res. 11:6925-6941.

Benne R, Van Den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC. 1986. Major transcript of the frame shifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell. 46:819-826.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A. 74:3171-3175.

Bhat GJ, Koslowsky DJ, Feagin JE, Smiley BL, Stuart K. 1990. An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. Cell. 61:885-894.

Bhattacharyya D, Diamond P, Basu S. 2015. An Independently folding RNA G-quadruplex domain directly recruits the 40S ribosomal subunit. Biochemistry. 54:1879-1885.

Biyani M, Nishigaki K. 2005. Structural characterization of ultra-stable higher-ordered aggregates generated by novel guanine-rich DNA sequences. Gene. 364:130-138.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem. 72:291-336.

Blaha GM, Polikanov YS, Steitz TA. 2012. Elements of ribosomal drug resistance and specificity. Curr Opin Struct Biol. 22:750-758.

Blom D, Burg Jv, Breek CK, Speijer D, Muijsers AO and Benne R. 2001. Cloning and characterization of two guide RNA-binding proteins from mitochondria of *Crithidia fasciculata*: gBP27, a novel protein, and gBP29, the orthologue of *Trypanosoma brucei* gBP21. Nucleic Acids Res. 29:2950-2962.

Blum B, Bakalara N, Simpson L. 1990. A model for RNA editing in kinetoplastid mitochondria: ''guide'' RNA molecules transcribed from maxicircle DNA provide the edited information. Cell. 60:189-198.

Blum B, Simpson L. 1990. Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. Cell. 62:391-397.

Böhm C, Katari VS, Brecht M, Göringer HU. 2012. *Trypanosoma brucei* 20 S editosomes have one RNA substrate-binding site and execute RNA unwinding activity. J Biol Chem. 287:26268-26277.

Brack C, Delain E, Riou G. 1972. Replicating, convalently closed, circular DNA from kinetoplasts of *Trypanosoma cruzi*. Proc Natl Acad Sci U S A. 69:1642-1646.

Breaker RR. 2011. Prospects for riboswitch discovery and analysis. Mol Cell. 43:867-879.

Brecht M, Niemann M, Schlüter E, Müller UF, Stuart K, Göringer HU. 2005. TbMP42, a protein component of the RNA editing complex in African trypanosomes has endo-exoribonuclease activity. Mol Cell. 17:621-630.

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. 1992. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell. 71:515-526.

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. 1992. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell. 71:527-542.

Carnes J, Trotter JR, Ernst NL, Steinberg A, Stuart K. 2005. An essential RNase III insertion editing endonuclease in *Trypanosoma brucei*. Proc Natl Acad Sci U S A. 102:16614-16619.

Carnes J, Stuart K. 2007. Uridine insertion/deletion editing activities. Meth Enzymol. 424:25-54.

Cech TR, Steitz JA. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 157:77-94.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell. 12:1-8.

Corell RA, Myler P, Stuart K. 1994. *Trypanosoma brucei* mitochondrial CR4 gene encodes an extensively edited mRNA with completely edited sequence only in bloodstream forms. Mol Biochem Parasitology. 64:65-74.

Corell RA, Read LK, Riley GR, Nellissery JK, Allen TE, Kable ML, Wachal MD, Seiwert SD, Myler PJ, Stuart KD. 1996. Complexes from *Trypanosoma brucei* that exhibit deletion editing and other editing-associated properties. Mol Cell Biol. 16:1410-1418.

Crick F. 1970. Central dogma of molecular biology. Nature. 227:561-563.

DasGupta S, Shelke SA, Li NS, Piccirilli JA. 2015. Spinach RNA aptamer detects lead(II) with high selectivity. Chem Commun (Camb). 51:9034-9037.

Davis JT, Spada GP. 2007. Supramolecular architectures generated by self-assembly of guanosine derivatives. Chem Soc Rev. 36:296-313.

De Lay NR, Garsin DA. 2016. The unmasking of 'junk' RNA reveals novel sRNAs: from processed RNA fragments to marooned riboswitches. Curr Opin Microbiol. 30:16-21.

Deng J, Schnaufer A, Salavati R, Stuart KD, Hol WG. 2004. High resolution crystal structure of a key editosome enzyme from *Trypanosoma brucei*: RNA editing ligase 1. J Mol Biol. 343:601-613.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 22:1775-1789.

Dey BK, Mueller AC, Dutta A. 2014. Long non-coding RNAs as emerging regulators of differentiation, development, and disease. Transcription. 5:e944014.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, *et al*. 2012. Landscape of transcription in human cells. Nature. 489:101-108.

Duncan CD, Weeks KM. 2010a. Nonhierarchical ribonucleoprotein assembly suggests a strain-propagation model for protein-facilitated RNA folding. Biochemistry. 49:5418-5425.

Duncan CD, Weeks KM. 2010b. The Mrs1 splicing factor binds the bI3 group I intron at each of two tetraloop-receptor motifs. PLoS One. 5:e8983.

Englund PT. 1978. The replication of kinetoplast DNA networks in *Crithidia fasciculata*. Cell. 14:157-168.

Eperon IC, Janssen JW, Hoeijmakers JH, Borst P. 1983. The major transcripts of the kinetoplast DNA of *Trypanosoma brucei* are very small ribosomal RNAs. Nucleic Acids Res. 11:105-125.

Ernst NL, Panicucci B, Igo RP Jr, Panigrahi AK, Salavati R and Stuart K. 2003. TbMP57 is a 3' terminal uridylyl transferase (TUTase) of the *Trypanosoma brucei* editosome. Mol Cell. 11:1525-1536.

Feagin JE, Jasmer DP, Stuart K. 1987. Developmentally regulated addition of nucleotides within apocytochrome b transcripts in *Trypanosoma brucei*. Cell. 49:337-345.

Feagin JE, Abraham JM, Stuart K. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell. 53:413-422.

Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, Stark H. 2015. Structure of the *E. coli* ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. Nature. 520:567-570.

Frech GC, Simpson L. 1996. Uridine insertion into preedited mRNA by a mitochondrial extract from *Leishmania tarentolae*: stereo-chemical evidence for the enzyme cascade model. Mol Cell Biol. 16:4584-4589.

Gao G, Simpson L. 2003. Is the *Trypanosoma brucei* REL1 RNA ligase specific for U-deletion RNA editing and is the REL2 RNA ligase specific for U-insertion editing? J Biol Chem. 278:27570-27574.

Gilbert W. 1978. Why genes in pieces? Nature. 271:501

Gilbert W. 1985. Genes-in-pieces revisited. Science. 228:823-824.

Gilbert W. 1986. Origin of life: The RNA world. Nature 319:618

Gluenz E, Povelones ML, Englund PT, Gull K. 2011. The kinetoplast duplication cycle in *Trypanosoma brucei* is orchestrated by cytoskeleton-mediated cell morphogenesis. Mol Cell Biol. 31:1012-1021.

Golas MM, Böhm C, Sander B, Effenberger K, Brecht M, Stark H, Göringer HU. 2009. Snapshots of the RNA editing machine in trypanosomes captured at different assembly stages *in vivo*. EMBO J. 28:766-778.

Göringer HU (ed). 2008. RNA Editing. Nucleic Acids and Molecular Biology. vol 20. Heidelberg: Springer Publishing.

Göringer HU. 2012. 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. Annu Rev Microbiol. 66:65-82.

Göringer HU, Koslowsky DJ, Morales TH, Stuart K. 1994. The formation of mitochondrial ribonucleoprotein complexes involving gRNA molecules in *Trypanosoma brucei*. Proc Natl Acad Sci U S A. 91:1776-1780.

Göringer HU, Katari VS, Böhm C. 2011. The structural landscape of native editosomes in African trypanosomes. Wiley Interdiscip Rev RNA. 2:395-407.

Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell. 35:849-857.

Guo JU, Agarwal V, Guo H, Bartel DP. 2014a. Expanded identification and characterization of mammalian circular RNAs. Genome Biol. 15:409.

Guo Y, Zhou L, Xu L, Zhou X, Hu J, Pei R. 2014b. Multiple types of logic gates based on a single G-quadruplex DNA strand. Sci Rep. 4:7315.

Grundy FJ, Rollins SM, Henkin TM. 1994. Interaction between the acceptor end of tRNA and the T box stimulates antitermination in the *Bacillus subtilis* tyrS gene: a new role for the discriminator base. J Bacteriol. 176:4518-4526.

Hans J, Hajduk SL, Madison-Antenucci S. 2007. RNA editing- associated protein 1 null mutant reveals link to mitochondrial RNA stability. RNA. 13:881-889.

Hashimi H, Zíkova A, Panigrahi AK, Stuart K, Lukeš J. 2008. TbRGG1, an essential protein involved in kinetoplastid RNA metabolism that is associated with a novel multiprotein complex. RNA. 14:970-980.

Hardin JW, Warnasooriya C, Kondo Y, Nagai K, Rueda D. 2015. Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. Nucleic Acids Res. 43:10963-10974.

Hayman ML, Read LK. 1999. *Trypanosoma brucei* RBP16 is a mitochondrial Y-box family protein with guide RNA binding activity. J Biol Chem. 274:12067-12074.

Hensgens LA, Brakenhoff J, De Vries BF, Sloof P, Tromp MC, Van Boom JH, Benne R. 1984. The sequence of the gene for cytochrome c oxidase subunit I, a frameshift containing gene for cytochrome c oxidase subunit II and seven unassigned reading frames in *Trypanosoma brucei* mitochondrial maxi-circle DNA. Nucleic Acids Res. 12:7327-7344.

Hermann T, Schmid B, Heumann H, Göringer HU. 1997 A three-dimensional working model for a guide RNA from *Trypanosoma brucei*. Nucleic Acids Res. 25:2311-2318.

Hoeijmakers JH, Weijers PJ. 1980. The segregation of kinetoplast DNA networks in *Trypanosoma brucei*. Plasmid. 4:97–116

Igo RP Jr, Palazzo SS, Burgess ML, Panigrahi AK, Stuart K. 2000. Uridylate addition and RNA ligation contribute to the specificity of kinetoplastid insertion RNA editing. Mol Cell Biol. 20:8447-8457.

Igo RP Jr, Weston DS, Ernst NL, Panigrahi AK, Salavati R, Stuart K. 2002. Role of uridylate-specific exoribonuclease activity in *Trypanosoma brucei* RNA editing. Eukaryot Cell. 1:112-118.

Ingram VM. 1957. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. Nature. 180:326-328.

Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. 3:318-356.

Jensen RE, Englund PT. 2012. Network news: the replication of kinetoplast DNA. Annu Rev Microbiol. 66:473-491.

Johansson J, Mandin P, Renzoni A, Chiaruttini C, Springer M, Cossart P. 2002. An RNA thermosensor controls expression of virulence genes in *Listeria* monocytogenes. Cell. 110:551-561.

Kable ML, Seiwert SD, Heidmann S, Stuart K. 1996. RNA editing: a mechanism for gRNA-specified uridylate insertion into pre-cursor mRNA. Science. 273:1189-1195.

Kang X, Rogers K, Gao G, Falick AM, Zhou S and Simpson L. 2005. Reconstitution of uridine-deletion precleaved RNA editing with two recombinant enzymes. Proc Natl Acad Sci U S A. 102:1017-1022.

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 316:1484-1488.

Katari VS, van Esdonk L, Göringer HU. 2013. Molecular crowding inhibits U-insertion/deletion RNA editing *in vitro*: consequences for the *in vivo* reaction. PLoS One. 8:e83796.

Klein DJ, Ferré-D'Amaré AR. 2006. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. Science. 313:1752-1756.

Kleisen MC, Borst P, Weijers PJ. 1976a. The structure of kinetoplast DNA. 1. The mini-circles of *Crithidia lucilae* are heterogeneous in base sequence. Eur J Biochem. 64:141-151.

Kleisen CM, Weislogel PO, Fonck K, Borst P. 1976b. The structure of kinetoplast DNA. 2. Characterization of a novel component of high complexity present in the kinetoplast DNA network of *Crithidia luciliae*. Eur J Biochem. 64:153-160.

Köller J, Nörskau G, Paul AS, Stuart K, Göringer HU. 1994. Different *Trypanosoma brucei* guide RNA molecules associate with an identical complement of mitochondrial proteins *in vitro*. Nucleic Acids Res. 22:1988-1995.

Köller J, Müller UF, Schmid B, Missel A, Kruft V, Stuart K, Göringer HU. 1997. *Trypanosoma brucei* gBP21: an arginine-rich mitochondrial protein that binds to guide RNA with high affinity. J Biol Chem. 272:3749-3757.

Koslowsky DJ, Bhat GJ, Perrollaz AL, Feagin JE, Stuart K. 1990. The MURF3 gene of *T. brucei* contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. Cell. 62:901-911.

Koslowsky D, Sun Y, Hindenach J, Theisen T, Lucas J. 2014. The insect-phase gRNA transcriptome in *Trypanosoma brucei*. Nucleic Acids Res. 42: 1873–1886.

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell. 31:147-157.

Kruse E, Voigt C, Leeder WM, Göringer HU. 2013. RNA helicases involved in U-insertion/deletion-type RNA editing. Biochim Biophys Acta. 1829:835-841.

Lander ES. 2016. The Heroes of CRISPR. Cell. 164:18-28.

Laurent M, Van Assel S, Steinert M. 1971. Kinetoplast DNA. A unique macromolecular structure of considerable size and mechanical resistance. Biochem Biophys Res Commun. 43:278-284.

Lukeš J, Guilbride DL, Votýpka J, Zíková A, Benne R, Englund PT. 2002. Kinetoplast DNA network: evolution of an improbable structure. Eukaryot Cell. 1:495-502.

Leung SS, Koslowsky DJ. 2001. Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: structure probing of an mRNA bound to is cognate gRNA. RNA. 7:1803-1816.

Liu J, Maxwell ES. 1990. Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. Nucleic Acids Res. 18:6565-6571.

Liu B, Liu Y, Motyka SA, Agbo EE, Englund PT. 2005. Fellowship of the rings: the replication of kinetoplast DNA. Trends Parasitol. 21:363-369.

Lu Z, Filonov GS, Noto JJ, Schmidt CA, Hatkevich TL, Wen Y, Jaffrey SR, Matera AG. 2015. Metazoan tRNA introns generate stable circular RNAs *in vivo*. RNA. 21:1554-1565.

Madison-Antenucci S, Sabatini RS, Pollard VW, Hajduk SL. 1998. Kinetoplastid RNA-editing-associated protein 1 (REAP-1): a novel editing complex protein with repetitive domains. EMBO J. 17:6368-6376.

Madison-Antenucci S, Hajduk SL. 2001. RNA editing associated protein 1 is an RNA binding protein with specificity for preedited mRNA. Mol Cell. 7:879-886.

Maslov DA, Simpson L. 1992. The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing. Cell. 70:459-467.

Maslov DA, Sturm NR, Niner BM, Gruszynski ES, Peris M, Simpson L. 1992. An intergenic G-rich region in *Leishmania tarentolae* kinetoplast maxicircle DNA is a pan-edited cryptogene encoding ribosomal protein S12. Mol Cell Biol. 12:56-67.

Maslov DA, Avila HA, Lake JA, Simpson L. 1994. Evolution of RNA editing in kinetoplastid protozoa. Nature. 368:345-348.

McGinnis JL, Weeks KM. 2014. Ribosome RNA assembly intermediates visualized in living cells. Biochemistry. 53:3237-3247.

McMahon M, Contreras A, Ruggero D. 2015. Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. Wiley Interdiscip Rev RNA. 6:173-189.

McManus MT, Shimamura M, Grams J, Hajduk SL. 2001. Identification of candidate mitochondrial RNA editing ligases from *Trypanosoma brucei*. RNA. 7:167-175.

Mellin JR, Cossart P. 2015. Unexpected versatility in bacterial riboswitches. Trends Genet. 31:150-156.

Miller MM, Read LK. 2003. *Trypanosoma brucei*: functions of RBP16 cold shock and RGG domains in macromolecular interactions. Exp Parasitol. 105:140-148.

Missel A, Souza AE, Nörskau G, Göringer HU. 1997. Gene disruption of a mitochondrial DEAD box protein in *Trypanosoma brucei* affects edited mRNAs. Mol Cell Biol. 17:4895-4903.

Miyamoto T, Razavi S, DeRose R, Inoue T. 2013. Synthesizing biomolecule-based Boolean logic gates. ACS Synth Biol. 2:72-82.

Müller UF. 2006. Re-creating an RNA world. Cell Mol Life Sci. 63:1278-1293.

Müller UF, Lambert L, Göringer HU. 2001 Annealing of RNA editing substrates facilitated by guide RNA binding protein gBP21. EMBO J. 20:1394-1404.

Müller UF, Göringer HU. 2002. Mechanism of the gBP21-mediated RNA/RNA annealing reaction: matchmaking and charge reduction. Nucleic Acids Res. 30:447-455.

Nam JW, Choi SW, You BH. 2016. Incredible RNA: Dual Functions of Coding and Noncoding. Mol Cells. 39:367-374.

Niemann M, Brecht M, Schlüter E, Weitzel K, Zacharias M, Göringer HU. 2008. TbMP42 is a structure sensitive ribonuclease that likely follows a metal ion catalysis mechanism. Nucleic Acids Res. 36:4465-4473.

Niemann M, Kaibel H, Schlüter E, Weitzel K, Brecht M, Göringer HU. 2009. Kinetoplastid RNA editing involves a 3' nucleotidyl phosphatase activity. Nucleic Acids Res. 37:1897-1906.

Novikova IV, Hennelly SP, Tung CS, Sanbonmatsu KY. 2013. Rise of the RNA machines: exploring the structure of long non-coding RNAs. J Mol Biol. 425:3731-3746.

Osman I, Tay ML, Pek JW. 2016. Stable intronic sequence RNAs (sisRNAs): a new layer of gene regulation. Cell Mol Life Sci. 2016 May 4. [Epub ahead of print]

Paige JS, Wu KY, Jaffrey SR. 2011. RNA mimics of green fluorescent protein. Science. 333:642-646.

Pandya-Jones A, Plath K. 2016. The "lnc" between 3D chromatin structure and X chromosome inactivation. Semin Cell Dev Biol. pii: S1084-9521:30087-8.

Payne M, Rothwell V, Jasmer DP, Feagin JE, Stuart K. 1985. Identification of mitochondrial genes in *Trypanosoma brucei* and homology to cytochrome c oxidase II in two different reading frames. Mol Biochem Parasitol. 15:159-170.

Pelletier M, Miller MM, Read LK. 2000. RNA-binding properties of the mitochondrial Y-box protein RBP16. Nucleic Acids Res. 28:1266-1275.

Pelletier M, Read LK. 2003. RBP16 is a multifunctional gene regulatory protein involved in editing and stabilization of specific mitochondrial mRNAs in *Trypanosoma brucei*. RNA. 9:457-468.

Peris M, Simpson AM, Grunstein J, Liliental JE, Frech GC, Simpson L. 1997. Native gel analysis of ribonucleoprotein complexes from a *Leishmania tarentolae* mitochondrial extract. Mol Biochem Parasitol. 85:9-24.

Pollard VW, Rohrer SP, Michelotti EF, Hancock K, Hajduk SL. 1990. Organization of minicircle genes for guide RNAs in *Trypanosoma brucei*. Cell. 63:783-790.

Pollard VW, Hajduk SL. 1991. *Trypanosoma equiperdum* minicircles encode three distinct primary transcripts which exhibit guide RNA characteristics. Mol Cell Biol. 11:1668-1675.

Pollard VW, Harris ME, Hajduk SL. 1992. Native mRNA editing complexes from *Trypanosoma brucei* mitochondria. EMBO J. 11:4429-4438.

Reifur L, Koslowsky DJ. 2008. *Trypanosoma brucei* ATPase subunit 6 mRNA bound to gA6-14 forms a conserved three-helical structure. RNA. 14:2195-2211.

Renger HC, Wolstenholme DR. 1972. The form and structure of kinetoplast DNA of *Crithidia*. J Cell Biol. 54:346-364.

Rhodes D, Lipps HJ. 2015. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 43:8627-8637.

Ribeiro MM, Teixeira GS, Martins L, Marques MR, de Souza AP, Line SR. 2015. G-quadruplex formation enhances splicing efficiency of PAX9 intron 1. Hum Genet. 134:37-44.

Rogers K, Gao G, Simpson L. 2007. Uridylate-specific 3' 5'-exoribonucleases involved in uridylate-deletion RNA editing in trypanosomatid mitochondria. J Biol Chem. 282:29073-29080.

Roy Chowdhury A, Bakshi R, Wang J, Yildirir G, Liu B, Pappas-Brown V, Tolun G, Griffith JD, Shapiro TA, Jensen RE, Englund PT. 2010. The killing of African trypanosomes by ethidium bromide. PLoS Pathog. 6:e1001226.

Rusché LN, Cruz-Reyes J, Piller KJ, Sollner-Webb B. 1997. Purification of a functional enzymatic editing complex from *Trypanosoma brucei* mitochondria. EMBO J. 16:4069-4081.

Rusché LN, Huang CE, Piller KJ, Hemann M, Wirtz E, Sollner-Webb B. 2001. The two RNA ligases of the *Trypanosoma brucei* RNA editing complex: cloning the essential band IV gene and identifying the band V gene. Mol Cell Biol. 21:979-989.

Sabatini R, Hajduk SL. 1995. RNA ligase and its involvement in guide RNA/mRNA chimera formation. Evidence for a cleavage-ligation mechanism of *Trypanosoma brucei* mRNA editing. J Biol Chem. 270:7233-7240.

Salavati R, Moshiri H, Kala S, Shateri Najafabadi H. 2011. Inhibitors of RNA editing as potential chemotherapeutics against trypanosomatid pathogens. Int J Parasitol Drugs Drug Resist. 2:36-46.

Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. 2012. Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. PLoS ONE. 7:e30733.

Sbicego S, Alfonzo JD, Estevez AM, Rubio MA, Kang X, Turck CW, Peris M, Simpson L. 2003. RBP38, a novel RNA-binding protein from trypanosomatid mitochondria, modulates RNA stability. Eukaryot Cell. 2:560-568.

Schmid B, Riley GR, Stuart K, Göringer HU. 1995. The secondary structure of guide RNA molecules from *Trypanosoma brucei*. Nucleic Acids Res. 23:3093-3102.

Schumacher MA, Karamooz E, Zíková A, Trantírek L, Lukeš J. 2006. Crystal structures of *T. brucei* MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. Cell. 126:701-711.

Seiwert SD, Stuart K. 1994. RNA editing: transfer of genetic information from gRNA to precursor mRNA *in vitro*. Science. 266:114-117.

Seiwert SD, Heidmann S, Stuart K. 1996. Direct visualization of uridylate deletion *in vitro* suggests a mechanism for kinetoplastid RNA editing. Cell. 84:831-841.

Sharp PA. 2009. The centrality of RNA. Cell. 136:577-580.

Shaw JM, Feagin JE, Stuart K, Simpson L. 1988. Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. Cell. 53:401-411.

Shay JW. 2016. Role of Telomeres and Telomerase in Aging and Cancer. Cancer Discov. [Epub ahead of print].

Simpson AG, Lukeš J, Roger AJ. 2002. The evolutionary history of kinetoplastids and their kinetoplasts. Mol Biol Evol. 19:2071-2083.

Sloof P, de Haan A, Eier W, van Iersel M, Boel E, van Steeg H, Benne R. 1992. The nucleotide sequence of the variable region in *Trypanosoma brucei* completes the sequence analysis of the maxicircle component of mitochondrial kinetoplast DNA. Mol Biochem Parasitol. 56:289-299.

Smola MJ, Calabrese JM, Weeks KM. 2015. Detection of RNA-Protein Interactions in Living Cells with SHAPE. Biochemistry. 54:6867-6875

Souza AE, Myler PJ, Stuart K. 1992. Maxicircle CR1 transcripts of *Trypanosoma brucei* are edited and developmentally regulated and encode a putative iron-sulfur protein homologous to an NADH dehydrogenase subunit. Mol Cell Biol. 12:2100-2107.

Steinert M, Van Assel S. 1975. Large circular mitochondrial DNA in *Crithidia luciliae*. Exp Cell Res. 96:406-409.

Steverding D. 2010. The development of drugs for treatment of sleeping sickness: a historical review. Parasit Vectors. 3:15.

Strobel EJ, Watters KE, Loughrey D, Lucks JB. 2016. RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. Curr Opin Biotechnol. 39:182-191.

Stuart K, Brun R, Croft S, Fairlamb A, Gürtler RE, McKerrow J, Reed S, Tarleton R. 2008. Kinetoplastids: related protozoan pathogens, different diseases. J Clin Invest. 118:1301-1310.

Sturm NR, Simpson L. 1990. Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. Cell. 61:879-884.

Sudarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, Breaker RR. 2006. Tandem riboswitch architectures exhibit complex gene control functions. Science. 314:300-304.

Sun M, Li W, Blomqvist K, Das S, Hashem Y, Dvorin JD, Frank J. 2015. Dynamical features of the *Plasmodium falciparum* ribosome during translation. Nucleic Acids Res. 43:10515-10524.

Tan R, Lan L. 2016. Guarding chromosomes from oxidative DNA damage to the very end. Acta Biochim Biophys Sin (Shanghai). [Epub ahead of print].

Trotter JR, Ernst NL, Carnes J, Panicucci B, Stuart K. 2005. A deletion site editing endo-nuclease in *Trypanosoma brucei*. Mol Cell. 20:403-412.

Vanhamme L, Perez-Morga D, Marchal C, Speijer D, Lambert L, Geuskens M, Alexandre S, Ismaïli N, Göringer U, Benne R, Pays E. 1998. *Trypanosoma brucei* TBRGG1, a mitochondrial oligo(U)-binding protein that co-localizes with an *in vitro* RNA editing activity. J Biol Chem. 273:21825-21833.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, *et al*. 2001. The sequence of the human genome. Science. 291:1304-1351.

Wahl MC, Lührmann R. 2015. SnapShot: Spliceosome Dynamics I. Cell. 161:1474-e1.

Wang F, Liu X, Willner I. 2014. DNA switches: from principles to applications. Angew Chem Int Ed Engl. 54:1098-1129.

Wang SK, Wu Y, Ou TM. 2015. RNA G-Quadruplex: The New Potential Targets for Therapy. Curr Top Med Chem. 15:1947-1956.

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 460:711-716.

Weeks KM, Cech TR. 1996. Assembly of a ribonucleoprotein catalyst by tertiary structure capture. Science. 271:345-348.

Win MN, Smolke CD. 2008. Higher-order cellular information processing with synthetic RNA devices. Science. 322:456-460

Yu LE, Koslowsky DJ. 2006. Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: structure probing of a gRNA bound to its cognate mRNA. RNA. 12:1050-1060.

Zappulla DC, Cech TR. 2004. Yeast telomerase RNA: a flexible scaffold for protein subunits. Proc Natl Acad Sci U S A. 101:10024-10029.

Zhang J, Ferré-D'Amaré AR. 2015. Structure and mechanism of the T-box riboswitches. Wiley Interdiscip Rev RNA. 6:419-433.

Research Aim

Thirty years ago Rob Benne and colleagues described the first form of RNA-editing within the mitochondrial pre-mRNAs of African trypanosomes. Since then, the basal reaction mechanism, the inventory and global structure of the catalytic editing machinery and several accessory factors have been identified. However, three decades of research provided no compelling evidence for either the biological necessity or the evolutionary origin of U-insertion/deletion RNA-editing. Furthermore, the physicochemical as well as structural details of the pre-mRNA/editosome interaction are not understood.

In **Chapter I**, I elucidate the 2D-structure of five mitochondrial mRNAs by selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) chemical probing in their free and editosome-bound states. The chemical mapping data are used to calculate the minimum free energy (MFE) 2D-structures of the different transcripts to track the editosome-mediated RNA unfolding upon interaction. I also present the first evidence for G-quadruplex (GQ)-folds as a structural feature of two extensively edited (pan-edited) mRNAs.

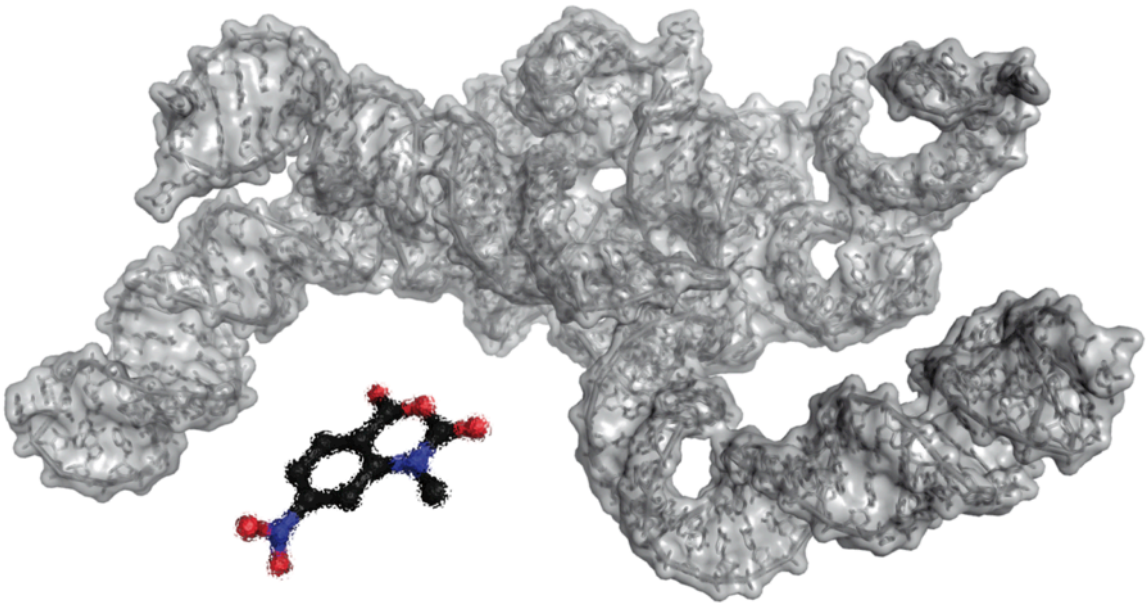
Chapter II focuses on a so far unrecognized structural feature of all pan-edited pre-mRNAs. I present bioinformatic and experimental evidence for multiple G-quadruplex folds exclusively occurring in pan-edited transcripts. In order to verify the existence of multiple GQ's I utilized a reverse transcriptase (RT)-stop assay to pinpoint the location of these stable structural elements. Furthermore, I analyze the fate of the GQ's as a consequence of RNA-editing and correlate the results with published data. This culminates in the first experimentally underpinned hypothesis regarding the evolutionary origin and necessity of RNA-editing in kinetoplastid organisms.

Mitochondria are the most crowded cellular compartment. **Chapter III** focuses on the question whether volume exclusion modulates pre-mRNA folding. I generate volume-occupied conditions *in vitro* by using a synthetic, inert crowding agent. In order to study the impact of the excluded volume effect on the 2D-structure and stability of the RPS12 transcript I utilize SHAPE chemical footprinting and UV-spectroscopy.

Chapter IV – U-insertion and deletion RNA-editing follows a multistep enzyme cascade mechanism. In order to acquire a deeper understanding of the underlying biophysical and biochemical determinants I screen for small molecular inhibitors using an established RNA-editing *in vitro* assay. I characterize the interaction of the most potent inhibitor with a minimalized RNA-editing substrate and the editosome by means of calorimetry, UV- and CD–spectroscopy, real time binding experiments as well as molecular modeling.

Chapter I

The RNA chaperone activity of the *Trypanosoma brucei* editosome raises the dynamic of bound pre-mRNAs



Abstract

Mitochondrial transcript maturation in African trypanosomes requires an RNA editing reaction that is characterized by the insertion and deletion of U-nucleotides into otherwise non-functional mRNAs. The reaction is catalyzed by editosomes and requires guide (g)RNAs as templates. Recent data demonstrate that the binding of pre-edited mRNAs to editosomes is followed by a chaperone-type RNA remodeling reaction. Here we map the changes in RNA folding using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE). We demonstrate that pre-mRNAs in their free state adopt intricately folded, highly stable 2D-structures. Editosome binding renders the pre-mRNAs to adopt 2D-conformations of reduced stabilities. On average about 30% of the nucleotides in every pre-mRNA are affected with a prevalence for U-nucleotides. The data demonstrate that the chaperone activity acts by increasing the flexibility of U-residues to lower their base-pairing probability. This results in a simplified RNA folding landscape with a reduced energy barrier to facilitate the binding of gRNAs. The data provide a first rational for the enigmatic U–specificity of the editing reaction.

Introduction

RNA chaperones catalyze the formation of the thermodynamically most stable RNA conformation by lowering the energetic barriers between misfolded RNA populations and by mediating the unwinding and refolding of non-functional RNA conformations (Herschlag, 1995; Rajkowitsch *et al*., 2007; Woodson, 2010; Grohman *et al*., 2013). As a consequence of that function, cells express a whole gamut of RNA chaperones, ATP-dependent RNA helicases and RNA annealing factors, which in effect contribute to all levels of gene expression, ribonucleoprotein assembly and RNA-mediated regulation. Within the different processes some RNA chaperones act as accessory factors, while others are integral components of the macromolecular machineries that catalyze the different reaction pathways (Russell, 2008). Mitochondrial transcript maturation in African trypanosomes requires an RNA editing reaction in which non-functional pre-mRNAs are converted into translatable mRNAs by the site-specific insertion and deletion of exclusively U-nucleotides (Aphasizhev & Aphasizheva, 2014). Some transcripts are edited by the insertion of literally hundreds of U's, which ultimately account for more than 50% of the mRNA sequence. The reaction is catalyzed by a 0.8MDa mitochondrial multienzyme complex termed the editosome (Göringer, 2012). RNA editing has been shown to involve RNA annealing factors (Müller & Göringer, 2002) as well as DEAD-box-type RNA helicases (Missel *et al*., 1997; Li *et al*., 2011; Kruse *et al*., 2013), however, recent evidence suggests that the editosome also executes an RNA chaperone function (Böhm *et al*., 2012). Although the molecular details of the chaperone reaction are not understood, it has been demonstrated that it is capable of refolding pre-edited mRNAs (Böhm *et al*., 2012), which are characterized by an unusual nucleotide bias. Especially all substantially *i.e.* pan-edited pre-mRNAs are extremely purine-rich with in some cases purine/pyrimidine (R/Y) ratios >2.5 (Supplementary Table 1).

By monitoring the local dynamic of >7200 nucleotides in several pre-edited mRNAs in their free and editosome-bound folding states we uncovered that the RNA chaperone activity of the *T. brucei* editosome acts by raising the dynamic of bound substrate RNAs. The reaction shows a preponderance for U-nucleotides and thus provides a first rational for the inexplicable U-specificity of the RNA editing reaction.

Results

Pre-edited mRNAs adopt thermodynamically highly stable 2D-structures. To map changes in the structural landscape of pre-edited mRNAs upon binding to editosomes we used the SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) chemical probing method developed by Weeks *et al*. (Merino *et al*. 2005; Low & Weeks; 2010; Weeks, 2015). SHAPE provides quantitative structural information with single-nucleotide resolution based on the accessibility of 2'-OH groups of conformationally flexible ribonucleotides to electrophilic acylation reagents. For the analysis we used 5 mitochondrial transcripts: the pre-mRNA of ribosomal protein S12 (RPS12), the transcript of subunit 3 of the NADH dehydrogenase (ND3), the pre-mRNAs of subunit 6 of the mitochondrial ATPase (A6) and of apocytochrome b (CYb) and the transcript of cytochrome oxidase I (COI). The different RNAs range in size from 282 nucleotides to 1647 nucleotides and include pre-mRNAs that are pan-edited (A6, ND3, RPS12), only marginally edited (CYb) and never-edited (COI). All RNAs were generated by run off *in vitro* transcription from linearized plasmid DNA constructs (Fig. 1A) and were refolded at native pH and ion conditions.

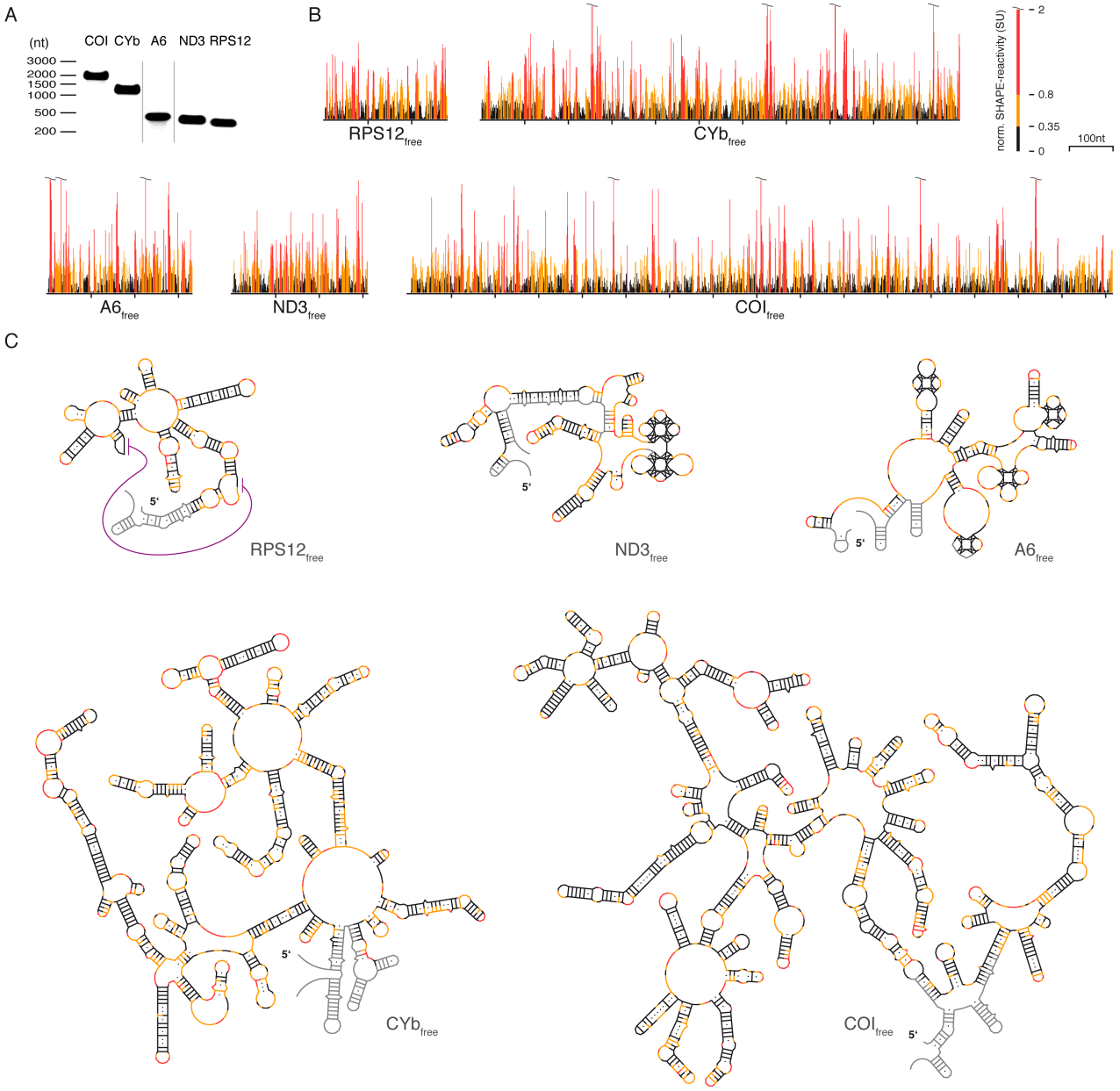


Figure 1. SHAPE-derived 2D-structures of *T. brucei* mitochondrial pre-mRNAs in their free folding states. (A) Gel electrophoretic characterization of the *T. brucei* COI-, CYb-, A6-, ND3- and RPS12-transcripts. (B) Normalized SHAPE-reactivity profiles of all 5 transcripts as free RNAs. Black: low (<0.35SU); yellow: medium (0.35≤SU<0.8); red: high (≥0.8SU) normalized SHAPE-reactivities. SU: SHAPE-unit. nt: nucleotides. A representative example of the individual steps to generate normalized SHAPE-profiles is given in Supplementary Fig. 2. (C) SHAPE-derived MFE-2D-structures of the 5 transcripts in their free folding states (coloring scheme as above). Purple line: pseudoknot fold in the RPS12-transcript. Grey: plasmid-derived 5'- and 3'-nt extensions or no data.

Structure probing was performed using the electrophile 1-methyl-7-nitroisatoic anhydride (1M7). 1M7 is a fast reacting compound with a half-life of 14sec (Mortimer & Weeks, 2007). The acylation is self-limiting due to hydrolysis of the reagent and has been shown to be insensitive to solvent accessibility constraints (McGinnis *et al.*, 2012).

Fig. 1B shows representative, normalized SHAPE-profiles for all 5 pre-mRNAs. Depending on the transcript up to 8 independent experiments were performed yielding Pearson correlation coefficients (r) of up to 0.87±0.04 (Supplementary Fig. 1). Altogether >3600 SHAPE-data points were generated. The individual reactivities were converted into pseudo Gibbs free energy terms (ΔG_{SHAPE}) to calculate experimentally derived minimal free energy (MFE) 2D-structures as shown

Table. 1. Summary of the thermodynamic folding characteristics of mitochondrial pre-mRNAs in their free and editosome-bound RNA folding state. ΔG-values are in kcal/mol. pe: pan-edited, me: marginally edited, ne: never edited.

		pre-mRNA				
		$RPS12_{pe}$	$ND3_{pe}$	$A6_{pe}$	CYb_{me}	COI_{ne}
in silico	$\Delta G_{RNAStructure}$	-102	-98	-98	-344	-468
	$\Delta G_{ViennaRNA}$	-103	-107	-107	-336	-459
RNA_{free}	ΔG_{SHAPE}	-152	-137	-137	-496	-823
	ΔG/nt	-0.45	-0.37	-0.34	-0.41	-0.48
	ΔG/bp	-1.46	-1.35	-1.59	-1.27	-1.47
	$r(nt_{bp}/nt_{ss})$	0.62	0.57	0.47	0.64	0.66
$RNA_{ed.-bound}$	ΔG_{SHAPE}	-127	-124	-113	-407	-725
	ΔG/nt	-0.38	-0.33	-0.28	-0.33	-0.43
	ΔG/bp	-1.39	-1.23	-1.62	-1.09	-1.39
	$r(nt_{bp}/nt_{ss})$	0.54	0.55	0.37	0.61	0.61
	ΔΔG	25	14	23	89	98
	ΔΔG/nt	0.08	0.04	0.06	0.07	0.06
	ΔΔG/bp	0.07	0.12	-0.03	0.17	0.08
	$\Delta nt_{basepaired}$ (%)	-7.7	-2.7	-10.1	-3.1	-4.5
	bp preserved (%)	84	82	50	55	52

in Fig. 1C. All pre-mRNAs adopt highly complex 2D-folds with ΔG's of −823kcal/mol (COI), −496kcal/mol (CYb), −152kcal/mol (RPS12), −137kcal/mol (ND3) and −137kcal/mol (A6). These values calculate to ΔG/nt-values between −0.34kcal/mol and −0.48kcal/mol and ΔG/bp-values between −1.27kcal/mol and −1.59kcal/mol (Table 1). In order to verify these values we performed temperature-dependent UV-hyperchromicity measurements and compared the measured hyperchromicities to the expected values of the SHAPE-derived 2D-structures (Supplementary Fig. 3). On average the two data sets differ only by 15% and thus corroborate the complex folds of the different transcripts. Importantly, nearest neighbor-based 2D-folding predictions (Hofacker, 2003; Reuter & Mathews, 2010) underestimate the experimentally derived thermodynamic stabilities on average by 30% (Table 1) emphasizing the need for structure probing experiments. Almost all known RNA 2D-structure features can be found in the 5 transcripts including a high number of G/U basepairs and a pseudoknot in the pre-mRNA of RPS12 (summarized in Supplementary Table 2).

Pre-edited A6 and ND3 transcripts contain multiple GQ-elements. One of the identified RNA folding features in the different pre-edited transcripts is of special interest: In two of the pre-mRNAs we identified multiple G-quadruplex (GQ)-folds (Millevoi *et al.*, 2012; Shrestha *et al.*, 2014). GQ-elements represents the most stable structures in DNA and RNA. The 2D-structure of the A6-transcript contains four and the ND3-transcript contains two GQ-folds (Fig. 1C). In order to derive an additional experimental verification of the GQ-elements we measured UV-thermal difference spectra (TDS) (Mergny *et al.*, 1998; Mergny *et al.*, 2005). TD-spectra represent the arithmetic

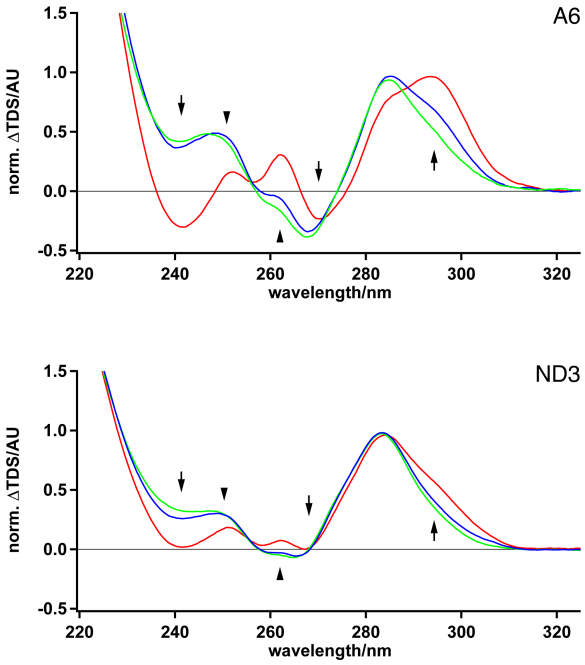


Figure 2. Experimental verification of GQ-folds - thermal difference (TD) spectra. Normalized double-difference TD-spectra of the A6- and ND3-transcript in the presence of KCl (red), NaCl (blue) and LiCl (green). Signature peaks of GQ-folds are marked by arrows (Mergny *et al.*, 2005). Additional differences are indicated by arrowheads. AU: arbitrary unit. Peak intensities decrease in the presence of GQ-destabilizing Na^+- and Li^+-ions.

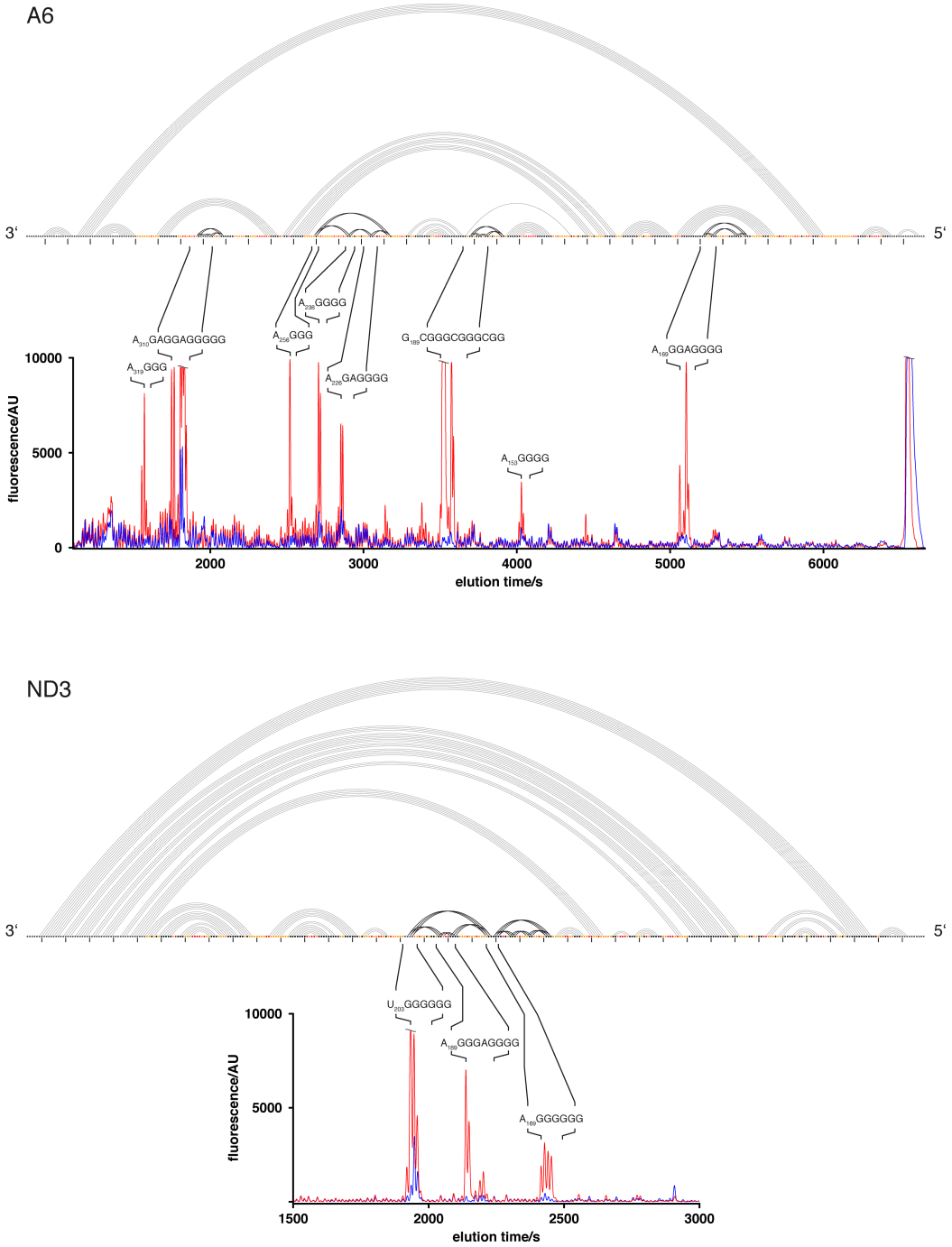


Figure 3. Experimental verification of GQ-folds – RT-stop assays. Capillary electrophoretic (CE)-traces of premature reverse transcriptase (RT)-stop assays of the A6- and ND3-transcripts in the presence of GQ-stabilizing K^+-ions (red) and GQ-destabilizing Na^+-ions (blue). Sequence identities of the individual RT-stop signals are shown on top of the different peaks. RNA secondary structures are annotated as arcs above the CE-traces. AU: arbitrary unit.

difference between a high- and low-temperature UV-absorbance spectrum *i.e* between RNA in its unfolded and fully folded states. In the presence of stabilizing K^+-ions, GQ-folds have been shown to display a set of signature peaks (243nm, 273nm and 295nm) (Mergny, 2005) and Fig. 2 shows representative TD-spectra for both, the A6 and ND3 transcripts. The two pre-mRNAs are characterized by all GQ-specific minima and maxima, which disappear in the presence of destabilizing Na^+- and Li^+-ions. The exact positions of the different GQ's were identified in premature reverse transcriptase

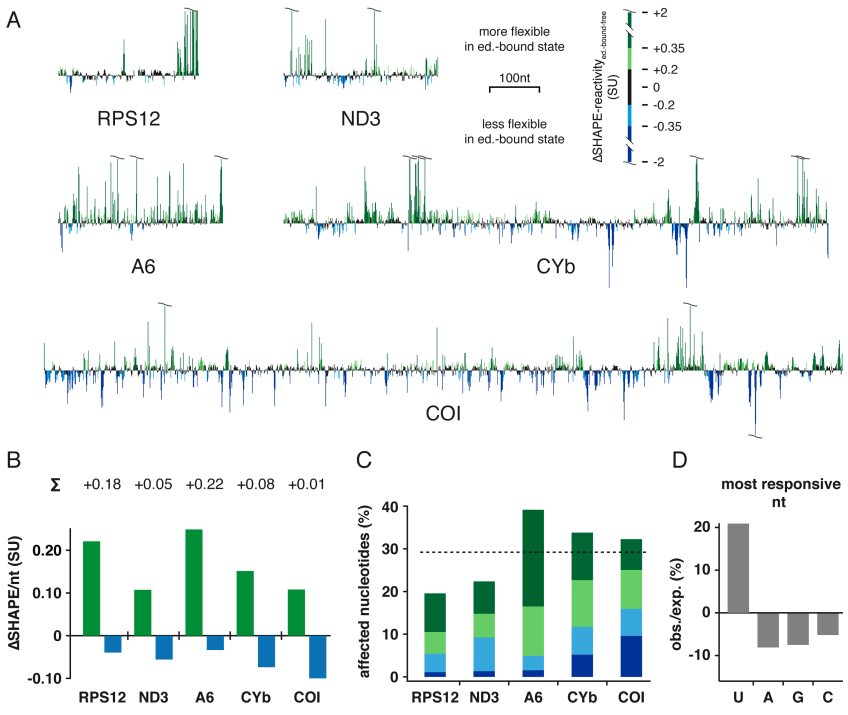


Figure 4. SHAPE-reactivity changes in the editosome-bound folding state. (**A**) Difference (Δ) SHAPE-reactivity profiles of the RPS12-, ND3-, A6-, CYb- and COI-transcripts. Green: nt-positions with increased SHAPE-reactivities in the editosome-bound folding state (dark green: +0.35≤SU≤+2; light green: +0.2≤SU<+0.35). Blue: nt-positions with decreased SHAPE-reactivities in the editosome-bound folding state (light blue: −0.2≥SU<-0.35; dark blue: −0.35≥SU≥−0.2. Black: non-responsive nt-positions (−0.2≥SU≤+0.2). (**B**) Bar graph summing up all increased (green) and decreased (blue) ΔSHAPE-reactivities. Numbers represent the difference of the two values and demonstrate a general increase in nt-flexibility upon editosome binding. SU: SHAPE-unit. nt: nucleotides. (**C**) Bar graph illustrating the percentage of affected nucleotides in all 5 transcripts. Stippled line: Arithmetic mean. (**D**) Bar graph illustrating the under-/over-representation of SHAPE-responsive nucleotides in the editosome-bound folding state. Values were calculated by determining the fraction of all 4 ribonucleotides (A, G, C, U) within the top 10% of responsive positions in the 5 transcripts divided by the relative abundance of each nt in that interval.

(RT) termination assays (Zheng *et al.*, 2014; Wu *et al.*, 2015). As shown in Fig. 3, at 75mM K^+ all G-nucleotide containing sequences involved in the formation of the individual GQ's were identified as strong RT-stop signals.

Editosome binding increases the dynamic of pre-edited mRNAs. Next we probed the structures of the selected pre-mRNAs in their editosome-bound folding state. For that we preincubated the transcripts with active editosomes in a 1:1 molar stoichiometry. Binding equilibrium was achieved at 27°C for 45-60min. Per RNA between 3 and 7 independent SHAPE-experiments were performed ultimately generating 3614 data points (*r* up to 0.81±0.07; Supplementary Fig. 1). As before, normalized SHAPE-reactivity profiles were plotted (Supplementary Fig. 4) to calculate experimentally-derived MFE-2D-structures for the 5 pre-mRNAs in their editosome-bound folding state. Importantly, in their bound form, all transcripts adopt less stable 2D-folds with on average 15% reduced Gibbs free energies (ΔΔG's) (Table 1). Based on the specificity of the acylation reagent this indicates a change in the folding landscape to increase the flexibility of the different pre-mRNAs. Fig. 4A shows ΔSHAPE-profiles of all pre-mRNAs highlighting the reactivity differences (increases/decreases) between the two folding states (free *vs.* editosome-bound). In spite of the different sizes of the transcripts the data demonstrate a global increase in reactivity especially for all edited transcripts. Reactivity differences generally spread throughout the entire primary sequences with individual nucleotides showing hyperreactive characteristics (>|2SU|) (McGinnis *et al.*, 2012). A quantitative comparison of the reactivity differences is given in Fig. 4B. While the never-edited COI transcript is only marginally affected (0.01SU), all edited transcripts (RPS12, A6, ND3, CYb) show net reactivity increases between 0.05-0.22SU. The same picture can be drawn from the arithmetic mean (AM) of the normalized SHAPE-reactivities (Supplementary Fig. 5). On average 29% of the nucleotide positions in every

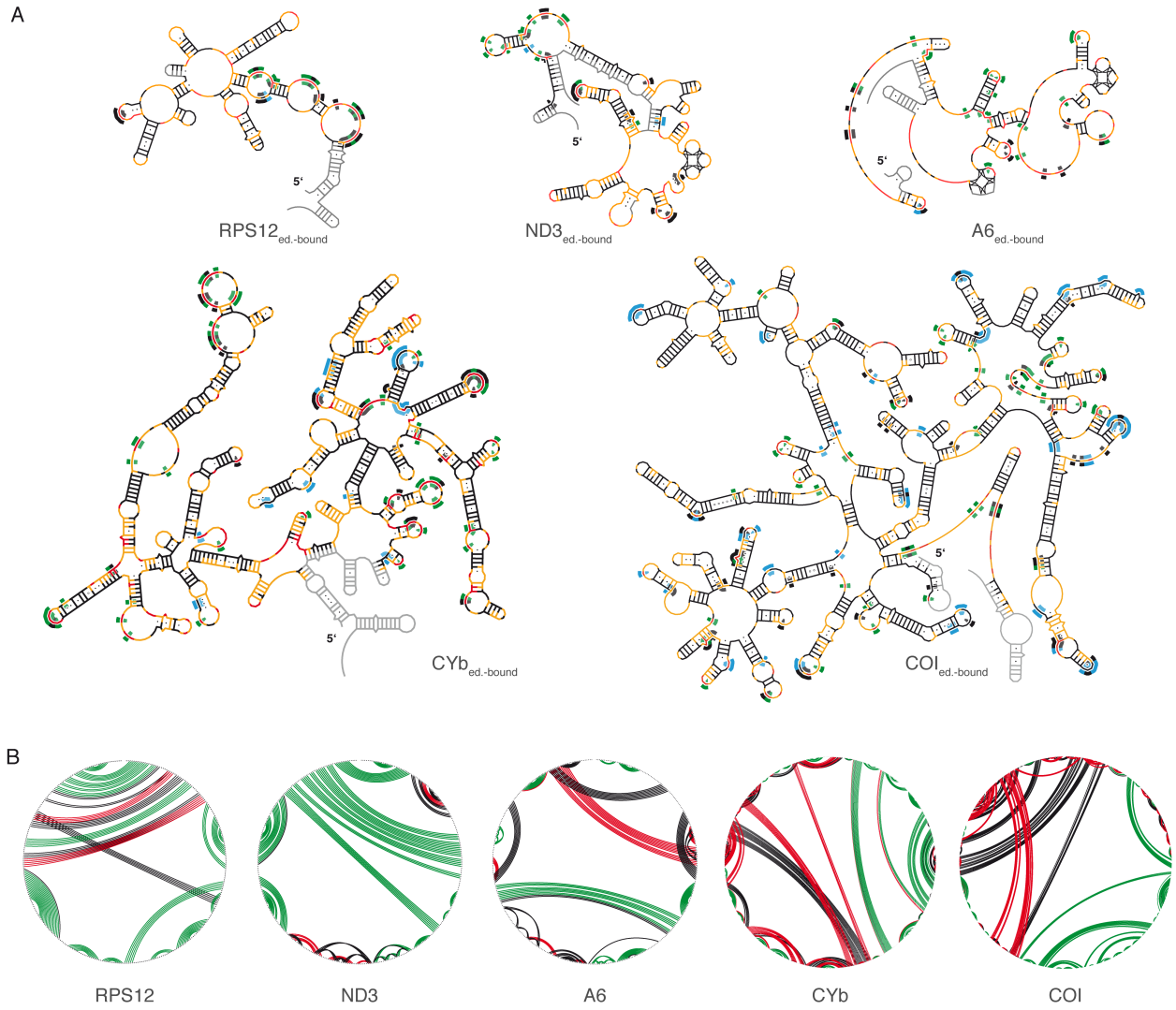


Figure 5. Mapping the structural transition from free RNA to editosome-bound RNA. (**A**) SHAPE-derived MFE-2D-structures of all 5 transcripts in their editosome-bound conformation. U-nt within the top 10% of responsive nt-positions are shown in green (increased SHAPE-reactivity) and blue (decreased SHAPE-reactivity). Black: non-uridine responsive nt. (**B**) Circle plot representations of the SHAPE-derived 2D-structures of the 5 pre-mRNAs. Base pairs are shown as colored lines. Green: basepairs present in both RNA folding states. Red: basepairs unique to the editosome-bound RNA conformations. Black: basepairs unique to the free RNA folding states. Plots were generated using CircleCompare of RNAstructure (Hajdin *et al.*, 2013).

pre-mRNA are affected in the conformational transition (Fig. 4C). Furthermore, by analyzing the correlation between nucleotide identity and changes in SHAPE-reactivity, we identified that U-nucleotides represent the most responsive nucleotides in the folding transition (Supplementary Fig. 6). From a total of 93 nucleotide positions that undergo a conversion from <0.35SU to >0.8SU 66 nucleotides or 71% are U-residues. Considering the nt-bias in the different pre-mRNAs and by plotting the ratio of observed/expected SHAPE-reactivities, the U-specificity of the chaperone activity becomes strikingly obvious. (Fig. 4D). This demonstrates that the RNA folding landscape of all 5 transcripts is simplified by increasing the flexibility of primarily U's.

Fig. 5A shows the SHAPE-derived MFE-2D-structures of the 5 RNAs in their editosome-bound form highlighting the 10% most responsive nucleotide positions. Compared to the free RNA folding states the average ΔG/nt-value of the 5 transcripts drops from −0.41kcal/mol to −0.33kcal/mol and the mean ΔG/bp-value

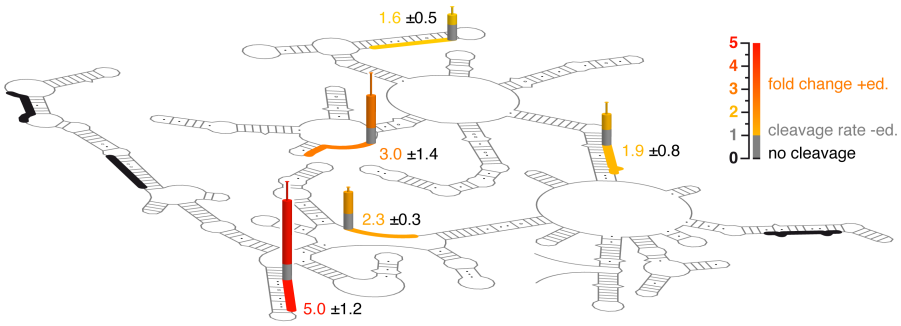


Figure 6. Guide RNA/pre-mRNA-hybrid formation in the free and editosome-bound RNA folding states. Background: SHAPE-derived 2D-fold of the CYb pre-mRNA. The binding sites of eight base-complementary DNA-oligonucleotides acting as quasi gRNAs are shown as thick lines. Colored bars indicate the change in RNaseH-based cleavage of the different DNA/pre-mRNA-hybrid molecules in the free and editosome-bound RNA folding states. Grey: RNaseH-cleavage in the free RNA folding state (=1). Yellow-red: >1-5-fold stimulus in the editosome-bound folding state. Black: no RNaseH cleavage. Errors are standard deviations (SD). A representative example of the data acquisition, normalization and quantification procedures is shown in Supplementary Fig. 8.

decreases from −1.43kcal/mol to −1.34kcal/mol (Table 1). Two GQ-elements in the A6 transcript and one GQ-fold in the ND3 transcript no longer exist demonstrating that the editosome chaperone activity possesses GQ-resolving potency analogous to the RecQ and Pif1-families of helicases (Millevoi *et al.*, 2012). Similarly, the pseudoknot in the RPS12 transcript is absent in the edititosome-bound folding state. Fig. 5B illustrates the 2D-folds and structural changes in a "circle plot" representation (Hajdin *et al.*, 2013). On average 65% of all basepairs in the free RNAs are preserved in the editosome-bound conformations. Roughly 20% of the basepairs are exclusive to either of the two folding states and the basepairing-content of all editosome-bound RNA structures is invariably lower than in the free RNA folding state (maximally 10%).

The editosome-bound folding state favors the formation of pre-mRNA/gRNA hybrid RNAs. The mechanistic rational for increasing the flexibility of editosome-bound RNAs is to overcome the structural rigidity of the different pre-edited transcripts to lower the energy barrier for subsequent steps in the editing cycle. Based on our current knowledge, the next step in the reaction is the annealing of a gRNA molecule to generate a pre-mRNA/gRNA hybrid, which defines the pre-mRNA endonucleolytic cleavage site for the U-insertion/deletion reaction. As a consequence of the data above, the formation of gRNA/pre-mRNA hybrid RNAs should be thermodynamically favored if the pre-mRNA is in its "flexible" *i.e.* editosome-bound state and disfavored if the transcript is in its "rigid" (free) configuration. To experimentally falsify this hypothesis we developed an assay in which the formation of pre-mRNA/gRNA hybrid is mimicked by short DNA-oligonucleotides as "guiding" molecules. This enabled us to quantitatively assess hybrid formation by RNaseH cleavage. Representative results are shown in Fig. 6 using CYb as a target pre-mRNA. Invariably, hybrid formation is favored if the transcript is in the editosome-bound folding state and not in the free RNA conformation. Depending on the guiding DNA-oligonucleotide up to 5-fold differences were measured between the two folding states, which demonstrates that at least one function of the editosome chaperone activity is to alter the pre-mRNA folding landscape to promote the formation of gRNA/pre-mRNA hybrid RNAs.

Discussion

Mitochondrial gene expression in African trypanosomes requires RNA editing. By inserting and deleting only U-nucleotides, the processing reaction converts non-functional, pre-edited

transcripts into translation-competent mRNAs. Editing follows a multistep catalytic reaction cycle that is mediated by a mitochondria-specific, high molecular mass protein particle known as the editosome (Göringer, 2012). Editosomes are characterized by a single RNA interaction domain on their surface (Böhm *et al*., 2012) and the binding of pre-edited mRNAs as substrates in the process has been shown to involve a chaperone-type RNA refolding reaction. Unfortunately, the biochemical and structural details of the chaperoning process are not understood (Böhm *et al*., 2012). Here we aimed at characterizing the editosome-inherent RNA chaperone activity by monitoring the folding transitions of 5 pre-edited mRNAs from their free folding state to their editosome-bound conformation using site-specific chemical probing (Merino *et al*., 2005; Low & Weeks, 2010; Weeks, 2015). We demonstrate that the pre-edited mRNAs in their free state adopt strikingly elaborate RNA folding characteristics with thermodynamic stabilities resembling structural RNAs. Two of the tested transcripts (A6, ND3) even contain multiple, up to 4 G-quadruplex elements, which represent the most stable folding motifs in DNA and RNA (Millevoi *et al*., 2012). However, editosome binding changes the structures of the different transcripts by simplifying their folding landscapes through an increase in RNA dynamic. About 30% of the nucleotide positions in every RNA are involved in the folding transition and invariably, the resulting editosome-bound MFE-structures have reduced thermodynamic stabilities. Although the different RNAs are not completely unfolded at the tested 1:1 editosome/pre-mRNA stoichiometry, fifty percent of the GQ-structures no longer exist in the editosome-bound RNA conformation. Importantly, the increase in RNA dynamic has a direct consequence for the editing reaction cycle and thus provides a rational for the chaperone function. Rate limiting step of the reaction is the binding of gRNA molecules to form short, intermolecular gRNA/pre-mRNA hybrid structures. These hybrid elements define the endonucleolytic cleavage sites within the pre-mRNA molecules, which represents the first step in the reaction cycle. The formation of gRNA/pre-mRNA hybrid RNAs has been shown to involve matchmaking-type RNA/RNA annealing factors (Müller & Göringer, 2002), however, the data above demonstrate that the editosome itself contributes to the formation of hybrid RNAs. By increasing the dynamic of bound pre-mRNAs, the activity raises the probability for a gRNA to anneal to its base-complementary target site, thereby initiating the reaction cycle.

Interestingly, the RNA folding landscape of the pre-edited mRNAs is altered by specifically increasing the dynamic of preferentially U-nucleotides in the different transcripts. This nucleotide specificity finds precedence in the RNA chaperone mechanism of the retroviral NC and hnRNP A1 proteins, which have been shown to execute a G-centric refolding activity (Grohman *et al*., 2013). U-nucleotides represent promiscuous base-pairing partners (Vendeix *et al*., 2009). They form homo- and hetero-basepairs with all other ribonucleotides and thus we propose that the U-specificity is a consequence of a selection process to globally lower the basepairing probability of pre-mRNAs by raising the dynamic of predominantly U's. This assures that the highest number of possible basepairs is affected. As a consequence, it is tempting to speculate that this might represent the driving force for the enigmatic U-specificity of the RNA editing process as a whole. The unfolding reaction should thermodynamically and perhaps kinetically "accelerate" as the editing reaction proceeds since the U-content dramatically increases (especially for all pan-edited mRNAs). Furthermore, the U-centricity might also explain why editing

involves roughly 10-times more U-insertions than U-deletions (Aphasizhev & Aphasizheva, 2014).

The data also suggest that the editosome does not discriminate between folded and misfolded RNAs or between transcripts that undergo editing or are never edited. COI-RNA as a never-edited transcript, behaved qualitatively and quantitatively identical to all other RNAs. This indicates that the difference between a productive and non-productive chaperone reaction is simply the probability increase in forming a gRNA/pre-mRNA hybrid, which drives the reaction towards the next step in the editing cycle. Which of the protein components of the editosome executes the chaperone activity is not known to date. It has been suggested that the activity resides within the different OB-fold proteins of the complex (Böhm *et al*., 2012) based on the observation that the yeast Rrp44 polypeptide executes an OB-fold-driven RNA chaperone activity (Lorentzen *et al*., 2008; Bonneau *et al*., 2009) Importantly, since 1M7 does not probe solvent accessibility (McGinnis *et al*., 2012) no conclusion as to the contact site(s) of the different transcripts with the editosome can be made. However, we predict that the catalytic complex displays a surface domain that is highly disordered and flexible (Czerwoniec *et al*., 2015), which enables it to interact with a diverse landscape of RNAs of different length and 3D-conformations. This flexibility is likely central to the chaperone activity because it allows the editosome to remain in contact with RNA as it is remodeled and edited (Wodson, 2010; Tompa & Csermely, 2004).

This is the first time that a dynamic reaction step of the RNA editing reaction cycle has been demonstrated experimentally. The editosome amplifies the intrinsic dynamic behavior of bound substrate RNAs (Dethoff *et al*., 2012) and induces multiple, U-specific RNA-equilibrium motions (Bokinsky *et al*., 2006; Shajani *et al*., 2007; Kim *et al*., 2014) to counteract the high thermodynamic stabilities of pre-edited mRNAs. This U-nucleotide centricity provides a first physicochemical rational for the perplexing U-specificity of the editing reaction.

Materials and methods

Cloning and RNA synthesis. Mitochondrial genes encoding subunit 6 of the mitochondrial ATPase (A6), apocytochrome b (CYb), NADH dehydrogenase subunit 3 (ND3), ribosomal protein S12 (RPS12) and cytochrome c oxidase subunit I (COI) were PCR-amplified from *T. brucei* Lister 427 genomic DNA (Cross, 1975) using the following DNA-oligonucleotide primers (KpnI and SacI restriction endonuclease recognition/cleavage sites are underlined): A6_forw. GGGGTACCAGAGGAA-TTTTGGGCGGAAGAG; A6_rev. CCGAGCTCCTATAAC-TCCAAAATCACAACTTTC; COI_forw. GGGGTACCA-TGTTTTTTCTATGTCTTGTGTGC; COI_rev. CCGAGC-TCTATATAAAAGAATAATAGGAAGG; CYb_forw. GG-GGTACCAGCGGAGAAAAAAGAAAGGGTC; CYb_rev. CCGAGCTCCTAATCTAACCTACACACTATG; ND3_forw. GGGGTACCTCAAAAAATCCTCGCCTTTT-TACTTTAG; ND3_rev. CCGAGCTCCTTGATGTTAGT-ATAAATGATTATATG; RPS12_forw. GGGGTACCCT-AATACACTTTTGATAACAAACTAAAG; RPS12_rev. CCGAGCTCCCTACCAAACATAAATGAACCTG. PCR-amplicons were cloned into the KpnI and SacI sites of phagemid pBS SKII$^-$ (Invitrogen) and transcripts were generated by run off *in vitro* transcription from linearized plasmids using T7-RNA polymerase. RNAs were purified from non-incorporated NTP's by size exclusion chromatography, EtOH-precipitated and dissolved in 10mMTris/HCl pH 7.5, 1mM EDTA (TE).

Editosome preparations. Editosomes were isolated from insect stage *Trypanosoma brucei* using the monomorphic strain Lister 427 (Cross, 1975).

Cells were grown in SDM79-medium (Brun & Schönenberger, 1979) to a cell density of $1x10^7$ cells/mL. After harvesting cells were disrupted by N_2-cavitation at isotonic conditions (Hauser *et al*., 1996). Mitochondrial vesicles were separated in Percoll step gradients and lysed in 1% (v/v) Nonidet P-40 in editing buffer (EB: 20mM HEPES/KOH, pH7.5, 30mM KCl, 10mM $Mg(OAc)_2$, 1mM DTT) containing 1mM PMSF, 1μg/mL leupeptin and 10μg/mL trypsin inhibitor. The detergent lysates were further separated by isokinetic ultracentrifugation in 10-35% (v/v) linear glycerol gradients as in Göringer *et al*., 1994 and after fractionation, 20S fractions were assayed for their RNA editing (Igo *et al*., 2000; Igo *et al*., 2002) (Supplementary Fig. 7) and RNA binding activity (Katari *et al*., 2013). RNA binding and RNA editing-competent fractions were stored at -20°C.

SHAPE-modification. The modification reagent 1-methyl-7-nitroisatoic anhydride (1M7) was synthesized as described (Turner *et al*., 2013). Mitochondrial pre-mRNAs (0.1μM) were denatured by heating to 95°C (2min) followed by snap cooling on ice. RNA refolding was achieved by equilibration in EB for 30min at 27°C, which represents the optimal growth temperature of insect-stage trypanosomes. RNA samples were split and treated either with 3.5mM 1M7 in DMSO or the same volume of neat DMSO. Modification reactions were quenched after 70sec by the addition of an equal volume of ddH_2O. RNAs were recovered by EtOH precipitation and desalted by size exclusion chromatography. 1M7-modifications in presence of 20S editosomes were performed at a molar pre-mRNA:editosome ratio of 1:1. Editosome/pre-mRNA complexes were incubated at 27°C for 45-60min and modified RNA was recovered by phenol extraction followed by EtOH precipitation and desalting by size exclusion chromatography.

Reverse transcription and data processing. Equimolar amounts of fluorescently labeled DNA oligonucleotide primers (see below) were annealed to 1M7-modified or unmodified RNA samples in 0.25xTE pH7.5 by heating to 95°C (2min), cooling to 50°C (10min) and snap cooling on ice. Reverse transcription was performed in 50mM Tris/HCl pH 8.3, 75mM KCl or NaCl (see below), 3mM $MgCl_2$, 2.5mM DTT, 0.25mM each dNTP and 0.75U/μL RiboLock RNase inhibitor (Invitrogen). The reaction was started by prewarming the samples for 1.5min prior to the addition of 5U/μL SuperScript III reverse transcriptase (RT, Invitrogen). The COI- and CYb-transcripts were incubated at 52°C for 1h. The pre-mRNA for RPS12 for 20min at 40°C and the ND3- and A6-transcripts for 5min at 50°C followed by 20min at 55°C. K^+ was used for the COI- and CYb-RNAs and Na^+ for the A6-, ND3- and RPS12-transcripts. Sequencing reactions were carried out using unmodified RNA, fluorescently labeled DNA oligonucleotide primer and ddNTPs at a final concentration of 0.125mM each. The ddNTP identity was chosen depending on the transcript analyzed. Reverse transcription was stopped by snap cooling and the addition of 0.1 vol. 4M NaOH followed by heating to 95°C (5min). Samples were pooled, EtOH precipitated and redissolved in HiDi formamid (ABI/Life technologies) for capillary electrophoresis. Raw electrophoretic traces were analyzed using SHAPEfinder (Vasa *et al*., 2008) utilizing the boxplot approach to determine the number of statistical outliers. Normalized SHAPE-reactivities were the result of averaging a minimum of 3 independent experiments. The following 5'-fluorescently labeled DNA-primer molecules were used: M13_reverse: 6-FAM/JOE/6-TAMRA-

CAGGAAACAGCTATGACCATG; T3 reverse: 6-FAM/JOE/TAMRA-AATTAACCCTCACTAAAGGGAAC; CYb_840-818: 6-FAM/JOE/6-TAMRA-CATAAACTT-ATCTGGGATTGCC; CYb_603-581: 6-FAM/JOE/6-TAMRA-CCTATCACAAAATGCATCAGAAC; CYb_310-282: 6-FAM/JOE/6-TAMRA-TAAATCATACTAATAT-ATGTGTGTCAAAC; COI_1391-1367: 6-FAM/JOE/6-TAMRA-GTCAACAAAAGCATACCATACAATG; COI_1144-1123: 6-FAM/JOE/6-TAMRA-CATAGTGG-AAATGTGCAACAAC; COI_1004-978: 6-FAM/JOE/6-TAMRA-CACATATCTGTGAATAAGAAGCTATAG; COI_877-857: 6-FAM/JOE/6-TAMRA-GATGCGCTCA-AACAAACATAC; COI_619-602: 6-FAM/JOE/6-TAMRA-CAGCCAAAACGGGAAGTG; COI_479-453: 6-FAM/JOE/6-TAMRA-CTAGAATACCAAGAAAATGCA-CAGAG.

SHAPE-directed RNA folding. Normalized SHAPE-reactivities were used as pseudo Gibbs free energy values to guide the folding of the different pre-mRNAs using RNAstructure v5.6 (Reuter & Mathews, 2010; Deigan *et al*., 2009). In addition, ShapeKnots (Hajdin *et al*., 2013) was used to search for pseudoknots using normalized SHAPE-reactivities and the default parameters of ShapeKnots (p1=0.35kcal/mol; p2=0.65kcal/mol). To extrapolate the ΔG of a pseudoknot-containing minimum free energy (MFE) structure to 27°C the ΔΔG of the MFE with and without the pseudoknot at 37°C was added to the free energy of the MFE at 27°C. G-quartet (GQ)-containing RNA structures were folded by initially forcing the corresponding sequence stretches into a single-stranded conformation. To estimate the ΔG of the complete structure, the free energy contribution of the GQ was calculated with the help of ViennaRNA v2.1.9 (Lorenz *et al*., 2013). In cases where the energy model of ViennaRNA was not applicable, the free energy of the competing, canonical 2D-structure was calculated using RNAstructure v5.6. Free energy contributions calculated by either method were finally added to the ΔG of the GQ-free structure. Non-canonical and canonical basepairs were introduced based on the primary data and isolated basepairs between highly reactive nt positions were removed from the final structures. Due to the unusual high distribution of SHAPE-reactivities in the A6-pre-mRNA in its editosome-bound state, box plot-normalized and averaged peak integrals corresponding to the modified reaction of both folding states were scaled against the invariable background followed by normalization to the outlier-revised free state (Low & Weeks, 2010).

Experimental verification of GQ-folds. GQ-structures were identified by premature RT-stop signals in the presence of GQ-stabilizing (K^+) and GQ-destabilizing (Na^+) ion conditions. Experiments were conducted analogous to the RT-SHAPE-procedure except that fluorescently labeled cDNAs were generated from unmodified transcripts in the presence of either 75mM K^+ or 75mM Na^+. cDNA fragments were resolved by capillary electrophoresis. Nucleotide identities were determined by cDNA-sequencing.

Thermal difference spectra. Spectroscopic measurements were carried out in a thermoelectrically controlled UV-spectro-photometer in 0.5mL quarts cuvettes. Prior to the analysis, RNAs were heat-denatured in 0.1mL TE pH7.5 (2min) and snap cooled on ice before the addition of concentrated folding buffer. Equilibration was allowed for 15min and RNA concentrations were set to an initial A_{260}=1. TD spectra were recorded in 5mM sodium cacodylate pH6.8 at 27°C and 95°C from 325nm to 220nm with a scan rate of 250nm/min generating 1 data point every 0.42nm. Monovalent ions (Li^+, Na^+, K^+) were

titrated from 1mM to 58mM using concentrated stock solutions keeping dilution effects to a minimum. Samples were allowed to equilibrate for 10min between each ion-titration step. TD spectra were generated by subtracting the normalized high temperature spectrum from the normalized low temperature spectrum. Double difference TD spectra were generated from the 1mM and 58mM cation titration spectra followed by normalization to the major peak around 290nm. All spectra were smoothed by Savitzky-Golay filtering (Savitzky & Golay, 1964).

UV-melting curves. Pre-mRNA transcripts were dissolved in 0.5xTE pH7.5 (50µL), heated to 95°C (2min) and snap cooled on ice before the addition of concentrated folding buffer to yield a final volume of 0.5mL and a final buffer concentration of 5mM Na cacodylate pH 6.8, 70 mM NaCl and 2mM $MgCl_2$. RNA concentrations were adjusted to A_{260}=0.5. Denaturation/renaturation profiles were measured at 260nm between 20°C and 95°C at a heating/cooling rate of 1°C/min (data aquisition: 0.3 data points/°C). Fraction folded (α)-plots were generated from the the melting curves as: $\alpha(T)=a/(a+b)$; a, b: distances to the upper and lower baselines. Baselines were set by linear regression. Hyperchromicity values (A_{rel}) were calculated by two different procedures. (1) "Nearest neighbor" approach using equation: A_{rel} (%) = $((A_{260}RNA_{unfolded}/(A_{260}ss + A_{260}ds))$ -1 x 100. Molar extinction coefficients of the fully unfolded RNAs, the double-stranded (ds) and the single-stranded (ss) subdomains of the SHAPE-derived 2D-structures were calculated using OligoCalc (Kibbe, 2007) followed by a conversion to absorbance values. GQ-forming nucleotides were treated as ss and non-canonical bp were treated as Watson/Crick bp. (2) "Fractional approach" using equation: A_{rel} (%) = $((1/f_{ss}+f_{ds}$ x N + f_{GQ} x M) −1) x 100. M=0.953 and N= 0.729 accounting for 37% hyperchromicity; f=fraction.

RNaseH-based "gDNA" annealing assay. The ability of free and editosome-bound pre-mRNAs to form gRNA/pre-mRNA hybrid RNAs was analyzed using a panel of "guiding" (g)DNA oligonucleotides complementary to different regions of the CYb-transcript (see numbers in brackets). The following oligodeoxynucleotides were used: CYb-1 (150-161): ACAAAATATCAA; CYb-2 (227-238): GTAAA-TATAATA; CYb-3 (255-266): ATTGACTTAAAT; CYb-4 (459-470): TCACTTCCCCAA; CYb-5 (651-662): AACATATCTCTC; CYb-7 (840-851): AAAAACA-AACCC; CYb-8 (967-978): ACTCATTCATAT; CYb-9 (1045-1056): CAAAAATAATAA. Annealed gDNA/pre-mRNA hybrid molecules were identified by RNaseH cleavage. For that, co-transcriptionally [^{32}P]-labeled CYb-RNA (1nM) was incubated with an eqimolar amount of editosomes and incubated for 15min at 27°C in EB. After the addition of 10-1000nM DNA-oligonucleotide RNaseH cleavage was performed for 20min at 27°C using 0.01U/µL *E. coli* RNaseH. RNA fragments were phenol-extracted followed by EtOH precipitation. Cleavage products were electrophoretically separated in 8M urea-containing 5% (w/v) polyacrylamide gels, visualized by phosphorimaging and densito-metrically quantified.

References

Aphasizhev R, Aphasizheva I. 2014. Mitochondrial RNA editing in trypanosomes: small RNAs in control. Biochimie. 100:125-131.

Böhm C, Katari VS, Brecht M, Göringer HU. 2012. *Trypanosoma brucei* 20S editosomes have one RNA substrate-binding site and execute RNA unwinding activity. J Biol Chem. 287:26268-26277.

Bokinsky G, Nivón LG, Liu S, Chai G, Hong M, Weeks KM, Zhuang X. 2006. Two distinct binding modes of a protein cofactor with its target RNA. J Mol Biol. 361:771-784.

Bonneau F, Basquin J, Ebert J, Lorentzen E, Conti E. 2009. The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. Cell. 139:547-559.

Brun R, Schönenberger M. 1979. Cultivation and *in vitro* cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. Acta Tropica. 36:289–292.

Cross GA. 1975. Identification, purification and properties of clone specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. Parasitology. 71:393–417.

Czerwoniec A, Kasprzak JM, Bytner P, Dobrychłop M, Bujnicki JM. 2015. Structure and intrinsic disorder of the proteins of the *Trypanosoma brucei* editosome. FEBS Lett. 589:2603-2610.

Deigan KE, Li TW, Mathews DH, Weeks KM. 2009. Accurate SHAPE-directed RNA structure determination. Proc Natl Acad Sci U S A. 106:97-102.

Dethoff EA, Chugh J, Mustoe AM, Al-Hashimi HM. 2012. Functional complexity and regulation through RNA dynamics. Nature. 482:322-330.

Göringer HU. 2012. 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. Annu Rev Microbiol. 66:65-82.

Göringer HU, Koslowsky DJ, Morales TH, Stuart K. 1994. The formation of mitochondrial ribonucleoprotein complexes involving guide RNA molecules in *Trypanosoma brucei*. Proc Natl Acad Sci U S A. 91:1776–1780.

Grohman JK, Gorelick RJ, Lickwar CR, Lieb JD, Bower BD, Znosko BM, Weeks KM. 2013. A guanosine-centric mechanism for RNA chaperone function. 2013. Science. 340:190-195.

Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. 2013. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proc Natl Acad Sci U S A. 110:5498-5503.

Hauser R, Pypaer M, Häusler T, Horn EK, Schneider A. 1996. *In vitro* import of proteins into mitochondria of *Trypanosoma brucei* and *Leishmania tarentolae*. J Cell Sci. 109:517-523.

Herschlag D. 1995. RNA chaperones and the RNA folding problem. J Biol Chem. 270:20871-20874.

Hofacker IL. 2003. Vienna RNA secondary structure server. Nucleic Acids Res. 31:3429-3431.

Igo RP Jr, Palazzo SS, Burgess ML, Panigrahi AK, Stuart K. 2000. Uridylate addition and RNA ligation contribute to the specificity of kinetoplastid insertion RNA editing. Mol Cell. Biol. 20:8447-8457.

Igo RP Jr, Weston DS, Ernst NL, Panigrahi AK, Salavati R, Stuart K. 2002. Role of uridylate-specific exoribonuclease activity in *Trypanosoma brucei* RNA editing. Eukaryot Cell. 1:112-118.

Katari VS, van Esdonk L, Göringer HU. 2013. Molecular crowding inhibits U-insertion/deletion RNA editing *in vitro*: consequences for the *in vivo* reaction. PLoS One. 8:e83796.

Kibbe WA. 2007. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res. 35 (Web Server issue):W43-46.

Kim H, Abeysirigunawarden SC, Chen K, Mayerle M, Ragunathan K, Luthey-Schulten Z, Ha T, Woodson SA. 2014. Protein-guided RNA dynamics during early ribosome assembly. Nature. 506:334-338.

Kruse E, Voigt C, Leeder WM, Göringer HU. 2013. RNA helicases involved in U-insertion/deletion-type RNA editing. Biochim Biophys Acta. 1829:835-841.

Li F, Herrera J, Zhou S, Maslov DA, Simpson L. 2011. Trypanosome REH1 is an RNA helicase involved with the 3'-5' polarity of multiple gRNA-guided uridine insertion/deletion RNA editing. Proc Natl Acad Sci U S A. 108:3542-3547.

Lorentzen E, Basquin J, Tomecki R, Dziembowski A, Conti E. 2008. Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. Mol Cell. 29:717-728.

Lorenz R, Bernhart SH, Qin J, Höner zu Siederdissen C, Tanzer A, Amman F, Hofacker IL, Stadler PF. 2013. 2D meets 4G: G-quadruplexes in RNA secondary structure prediction. IEEE/ACM Trans Comput Biol Bioinform. 10:832-844.

Low JT, Weeks KM. 2010. SHAPE-directed RNA secondary structure prediction. Methods. 52:150-158.

McGinnis JL. Dunkle JA, Cate JH, Weeks KM. 2012. The mechanisms of RNA SHAPE chemistry. J Am Chem Soc. 134:6617-6624.

Mergny JL, Phan AT, Lacroix L. 1998. Following G-quartet formation by UV-spectroscopy. FEBS Lett. 435:74-78.

Mergny JL, Li J, Lacroix L, Amrane S, Chaires JB. 2005. Thermal difference spectra: a specific signature for nucleic acid structures. Nucleic Acids Res. 33:e138.

Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc. 127:4223-4231.

Millevoi S, Moine H, Vagner S. 2012. G-quadruplexes in RNA biology. Wiley Interdiscip Rev RNA. 3:495–507.

Missel A, Souza AE, Nörskau G, Göringer HU. 1997. Disruption of a gene encoding a novel mitochondrial DEAD-box protein in *Trypanosoma brucei* affects edited mRNAs. Mol Cell Biol. 17:4895-4903.

Mortimer SA, Weeks KM. 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J Am Chem Soc. 129:4144-4145.

Müller UF, Göringer HU. 2002. Mechanism of the gBP21-mediated RNA/RNA annealing reaction: matchmaking and charge reduction. Nucleic Acids Res. 30:447-455.

Rajkowitsch L, Chen D, Stampfl S, Semrad K, Waldsich C, Mayer O, Jantsch MF, Konrat R, Bläsi U, Schroeder R. 2007. RNA chaperones, RNA annealers and RNA helicases. RNA Biol. 4:118-130.

Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics. 11:129.

Russell R. 2008. RNA misfolding and the action of chaperones. Front Biosci. 13:1-20.

Savitzky A, Golay MJE. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem. 36:1627–1639.

Shajani Z, Drobny G, Varani G. 2007. Binding of U1A protein changes RNA dynamics as observed by ^{13}C NMR relaxation studies. Biochemistry. 46:5875-5883.

Shrestha P, Xiao S, Dhakal S, Tan Z, Mao H. 2014. Nascent RNA transcripts facilitate the formation of G-quadruplexes. Nucleic Acids Res. 42:7236-7246.

Tompa P, Csermely P. 2004. The role of structural disorder in the function of RNA and protein chaperones. FASEB J. 18:1169-1175.

Turner R, Shefer K, Ares M Jr. 2013. Safer one-pot synthesis of the 'SHAPE' reagent 1-methyl-7-nitroisatoic anhydride (1m7). RNA. 19:1857-1863.

Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. 2008. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA. 14:1979-1990.

Vendeix FA, Munoz AM, Agris PF. 2009. Free energy calculation of modified base-pair formation in explicit solvent: A predictive model. RNA. 15:2278-2287.

Weeks KM. 2015. Toward all RNA structures, concisely. Biopolymers. 103:438-448.

Woodson SA. 2010. Taming free energy landscapes with RNA chaperones. RNA Biol. 7:677-686.

Wu RY, Zheng KW, Zhang JY, Hao YH, Tan Z. 2015. Formation of DNA:RNA hybrid G-quadruplex in bacterial cells and its dominance over the intramolecular DNA G-quadruplex in mediating transcription termination. Angew. Chem. Int. Ed. Engl. 54:2447-2451.

Zheng KW, Wu RY, He YD, Xiao S, Zhang JY, Liu JQ, Hao YH, Tan Z. 2014. A competitive formation of DNA:RNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site. Nucleic Acids Res. 42:10832-10844.

Author Contributions:
Conceived and designed experiments: HUG, WML
Performed experiments: WML, CV (gDNA assay), MB (cloning of pre-mRNA genes)

Supplementary Tables/Figures

Supplementary Table. 1. Nucleotide composition of the *T. brucei* mitochondrial pre-mRNAs used in this study.

	pre-mRNA				
	$RPS12_{pe}$	$ND3_{pe}$	$A6_{pe}$	CYb_{me}	COI_{ne}
length (nt)	282	322	344	1080	1647
A	85	110	117	327	372
U	93	100	75	505	788
G	75	86	132	179	285
C	29	26	20	69	202
R/Y	1.3	1.6	2.6	0.9	0.7

pe: pan-edited, me: marginally edited, ne: never edited.

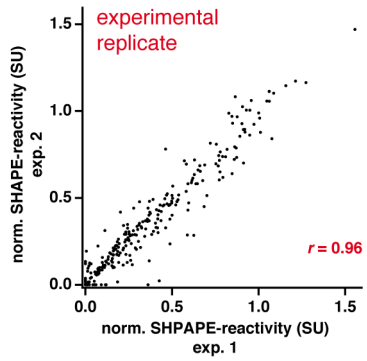
Supplementary Table. 2. Summary of the identified 2D-structural features of the 5 *T. brucei* mitochondrial pre-mRNAs in their free RNA folding states.

	pre-mRNA				
	$RPS12_{pe}$	$ND3_{pe}$	$A6_{pe}$	CYb_{me}	COI_{ne}
longest ds	17	10	9	17	15
longest ss	9	8	18	17	13
GU bp	21	15	17	71	107
other non-canonical bp	1	0	0	12	29
G-quadruplex	0	2	4	0	0
pseudoknots	1	0	0	0	0
hairpin loops	8	9	7	24	36
tri/tetra loops	3	5	2	7	14
internal loops	7	6	2	25	23
symmetric	2	2	1	8	8
asymmetric	5	4	1	17	15
external loops	1	1	1	1	1
multi-loops	2	2	3	8	11
bulges	2	4	3	12	26

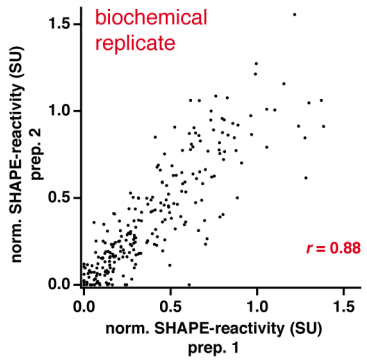
pe: pan-edited, me: marginally edited, ne: never edited.

Supplementary Figure 1. Experimental variation of SHAPE-modification experiments. (**A**) Representative Pearson plot of SHAPE-reactivity data of an experimental replicate (exp. 1, exp. 2) of the RPS12 transcript. (**B**) Representative Pearson plot of a biochemical replicate (prep. 1, prep. 2) of the RPS12 pre-mRNA. (**C**) Summary of Pearson correlation coefficients (*r*) and number of experimental replicates for all *T. brucei* transcripts in their free and editosome-bound (ed.-bound) folding states.

A



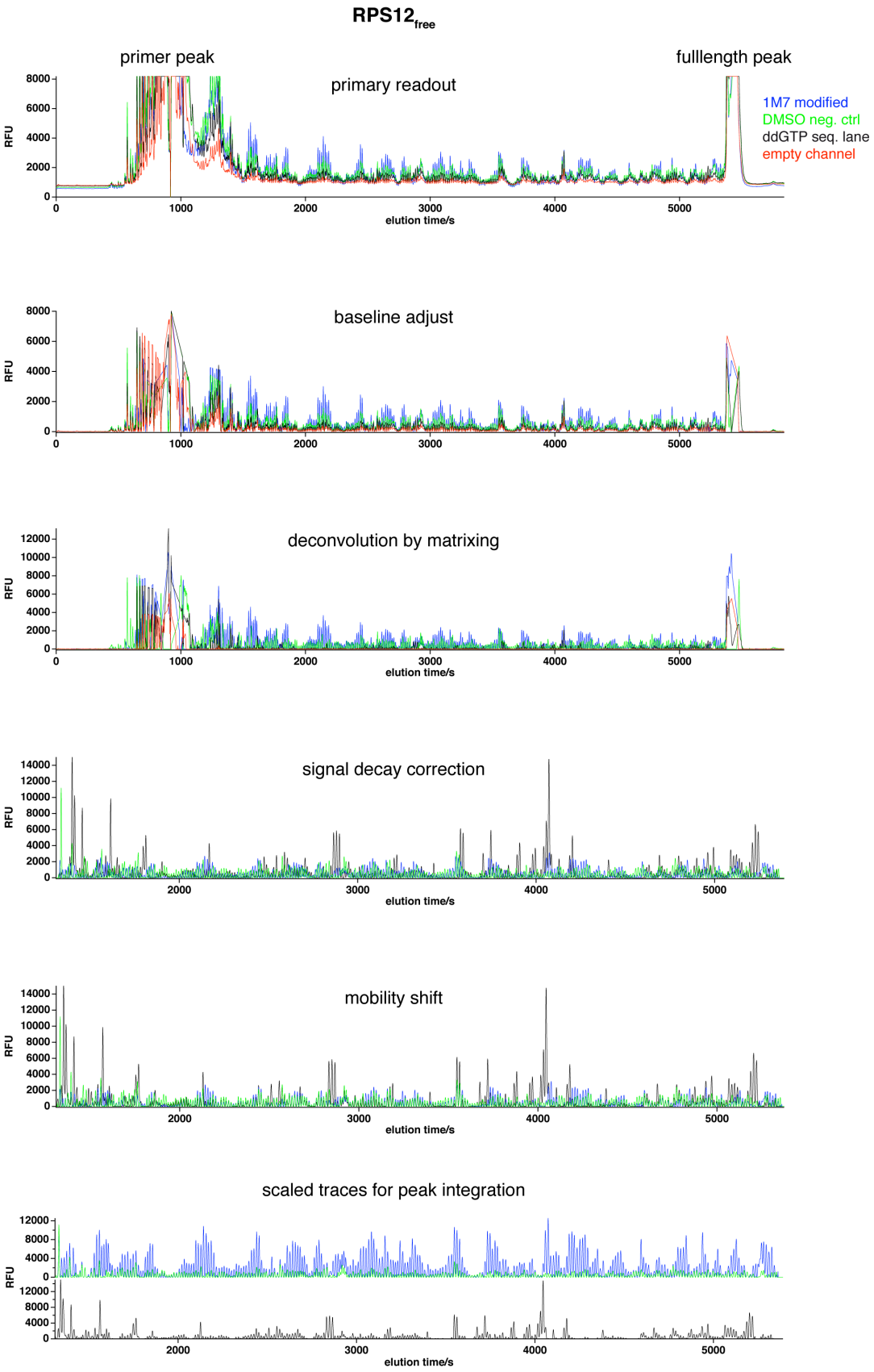
B



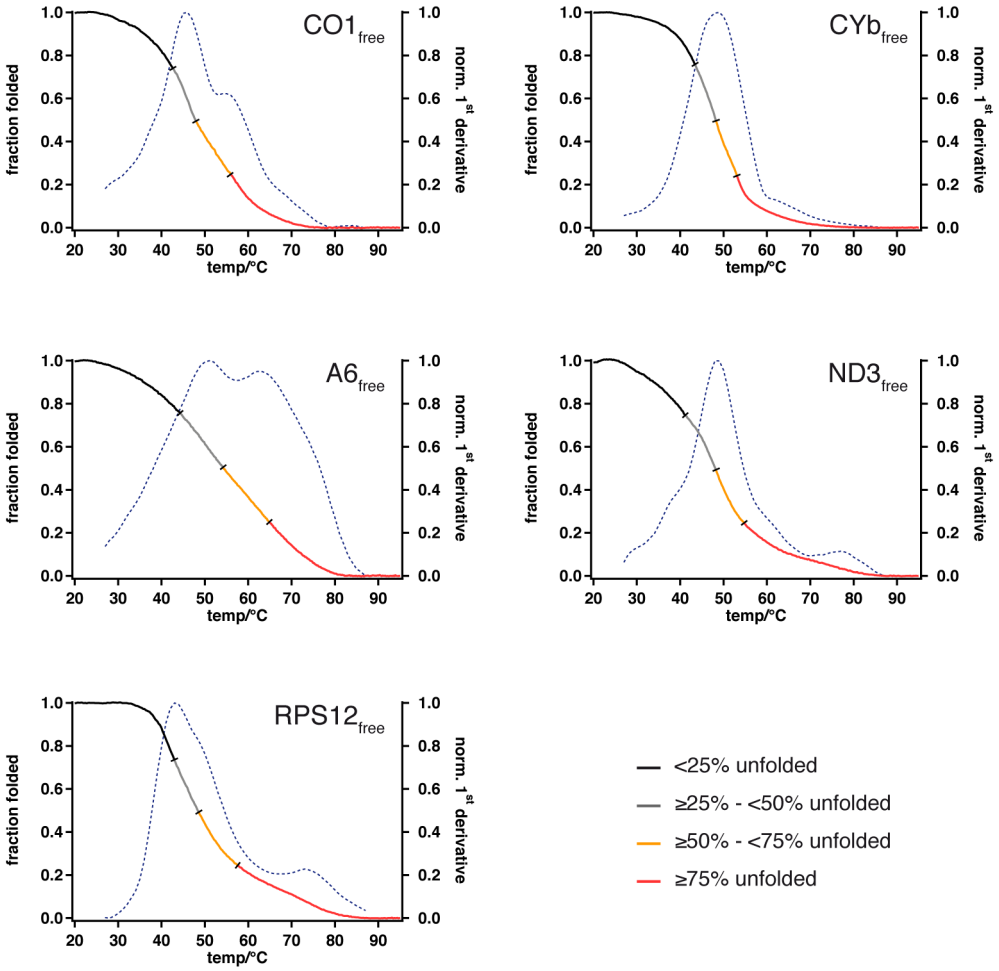
C

pre-mRNA	pearson (*r*) ±SD	replicates
$RPS12_{free}$	0.87 ± 0.04	5
$RPS12_{ed.-bound}$	0.81 ± 0.07	3
$ND3_{free}$	0.86 ± 0.06	4
$ND3_{ed.-bound}$	0.73 ± 0.10	7
$A6_{free}$	0.83 ± 0.07	5
$A6_{ed.-bound}$	0.79 ± 0.09	4
CYb_{free}	0.82 ± 0.06	8
$CYb_{ed.-bound}$	0.79 ± 0.07	3
COI_{free}	0.83 ± 0.05	3
$COI_{ed.-bound}$	0.60 ± 0.05	4

Supplementary Figure 2. Workflow to generate normalized SHAPE-profiles. Capillary electrophoresis (CE)-trace of the RPS12 transcript in its free folding state ($RPS12_{free}$) followed by all operational steps to generate normalized SHAPE-modification profiles. RFU: relative fluorescence unit.

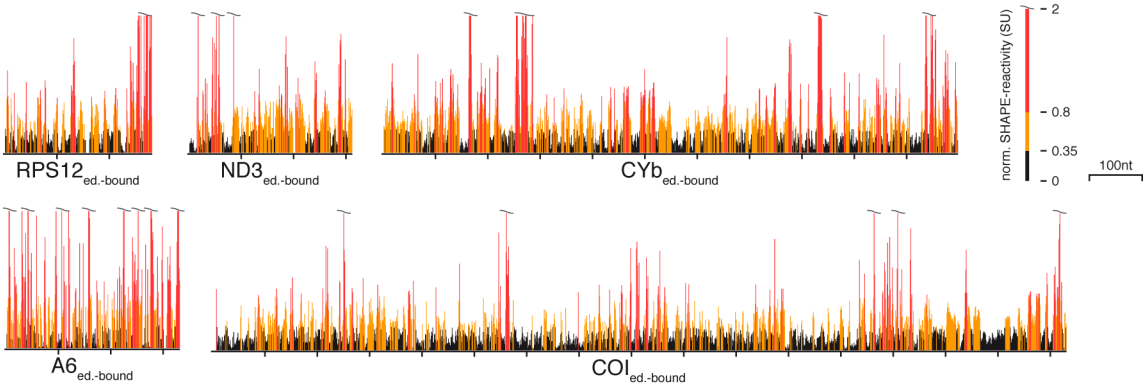


Supplementary Figure 3. UV-melting curves. Fraction folded (α) *versus* temperature plots (solid lines) and normalized 1st-derivatives (dashed lines) of the never-edited COI-RNA and the pre-edited CYb-, RPS12-, ND3- and A6-transcripts in their free folding states. Colors annotate the percentage of unfolded RNA (see legend). All transcripts show complex melting profiles with a main melting transition around 50°C and additional helix/coil transitions ≥70°C (A6, ND3, RPS12). Bottom table: Comparison of experimentally- and theoretically-derived hyperchromicities ($A_{rel.}$) (see Methods section). Errors are standard deviations (SD).

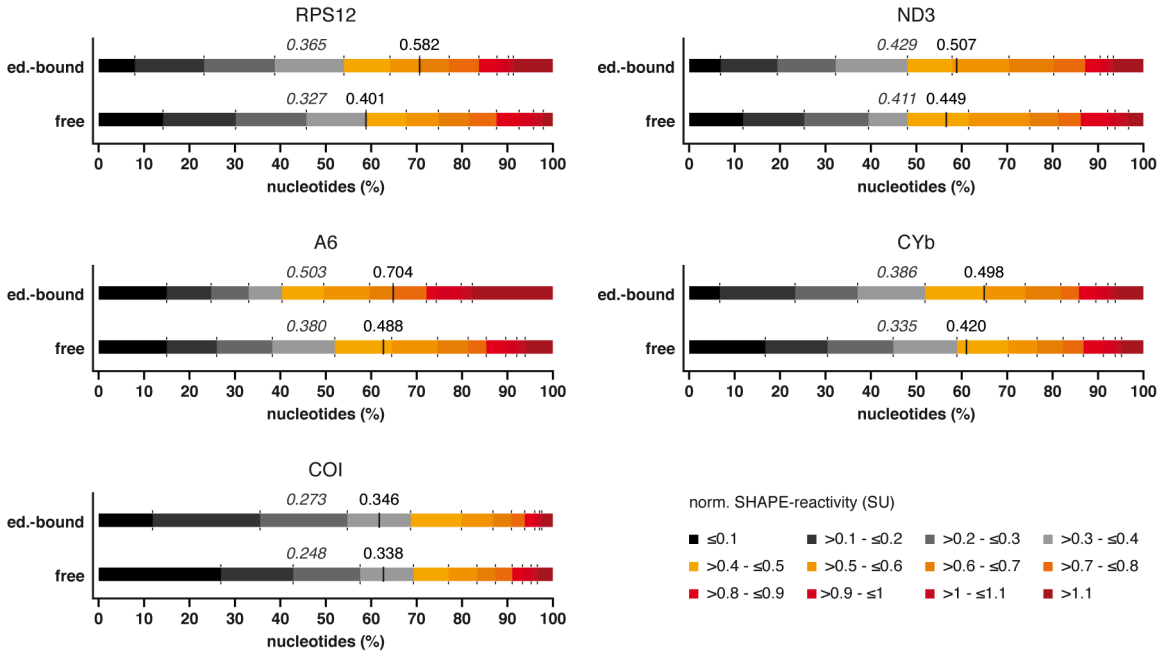


	measured hyperchromicity (%)	predicted hyperchromicity (%)	
pre-mRNA	UV-melting curves	"nearest neighbour"	"fractional approach"
COI_{free}	20.3±1.0	30	22
Cyb_{free}	31.5±1.6	32	21
$A6_{free}$	20.3±1.0	19	15
$ND3_{free}$	17.2±0.9	20	16
$RPS12_{free}$	20.5±1.2	25	20

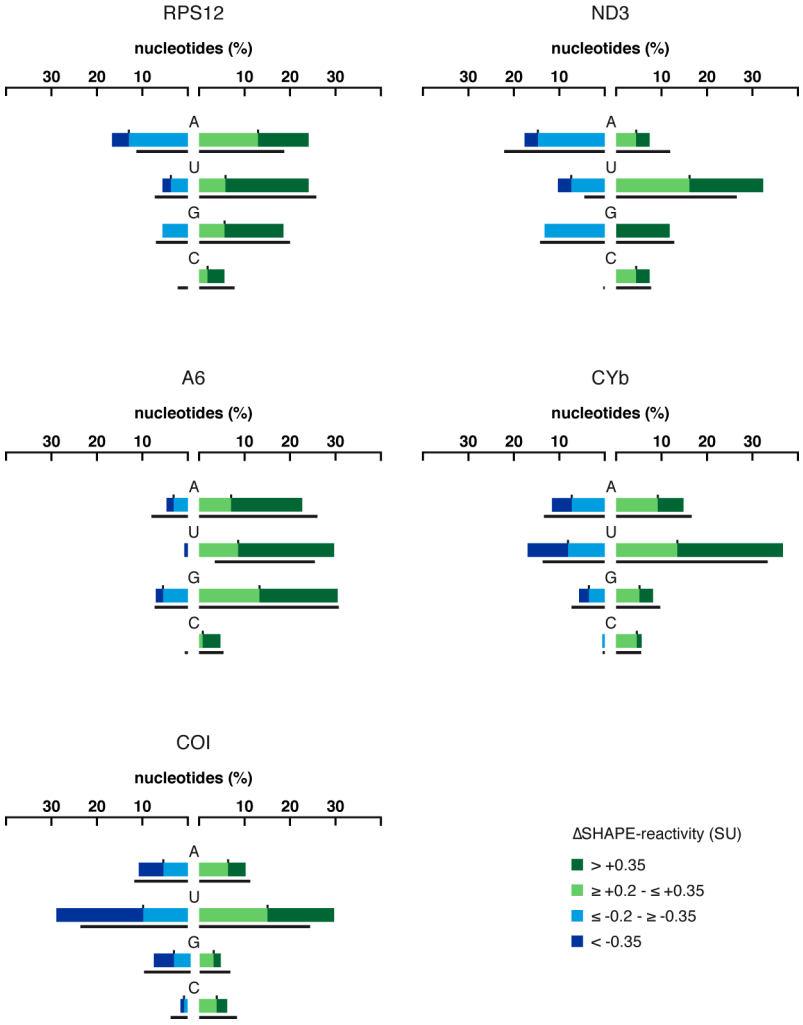
Supplementary Figure 4. SHAPE-reactivity profiles of *T. brucei* mitochondrial pre-mRNAs in their editosome-bound folding state. Normalized SHAPE-reactivity profiles of all 5 transcripts in their editosome-bound folding state (ed.-bound). Black: low (≤0.35SU); yellow: medium (0.35<SU≤0.8); red: high (>0.8SU) normalized SHAPE-reactivities. SU: SHAPE-unit. nt: nucleotides.



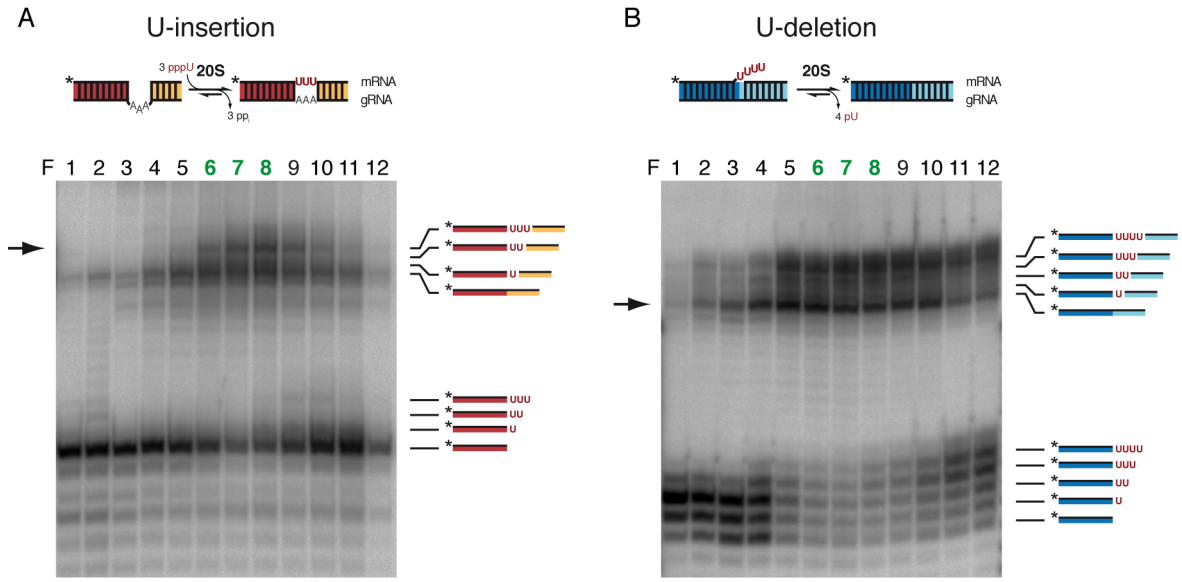
Supplementary Figure 5. Summary of SHAPE-reactivities in the free and editosome-bound RNA folding state. Histograms comparing the normalized SHAPE-reactivities of all 5 *T. brucei* mitochondrial transcripts in their free and editosome-bound (ed.-bound) folding state. SHAPE-reactivities were binned in 0.1 SHAPE-units (SU) and plotted as stacked bars in percent. The median SHAPE-reactivity is shown in italics. The arithmetic mean (AM) is marked as a black line. Editosome binding increases the AM of all transcripts especially the 3 pan-edited pre-mRNAs (A6, ND3, RPS12) indicating an increase in structural flexibility



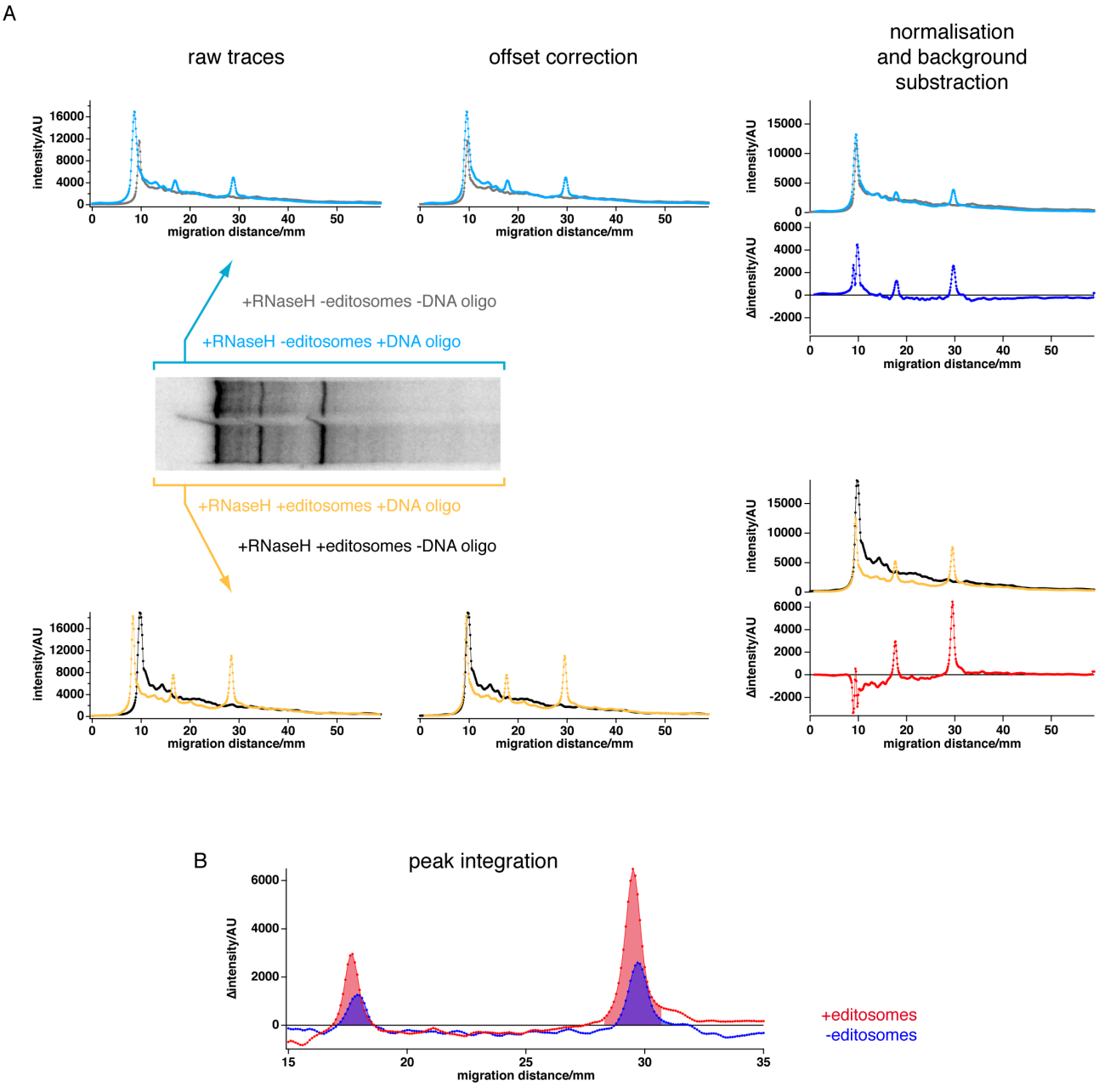
Supplementary Figure 6. Nucleotide specificity of the free to editosome-bound RNA-folding transition. For each pre-mRNA transcript, the number of affected nt was set to 100% and divided into 4 classes: nucleotides that become slightly more flexible (ΔSHAPE 0.2≤SU<0.35 - light green), significantly more flexible (ΔSHAPE ≥0.35SU - dark green), slightly less flexible (ΔSHAPE -0.2≥SU>-0.35 - cyan) and significantly less flexible (ΔSHAPE ≤-0.35 - dark blue) in the editosome-bound RNA folding state. A change of 0.35 in either direction can represent the transition from a bp- to a ss-conformation or *vice versa*. Data were plotted as stacked bars in percent. Black lines: relative abundance of the different nucleotides in the various transcripts



Supplementary Figure 7. *In vitro* RNA editing. Pre-cleaved *in vitro* U-insertion (A) and U-deletion (B) RNA editing assays to determine the editing activity of twelve *T. brucei* mitochondrial fractions (F1-12). The U-insertion assay monitors the insertion of 3 U nucleotides; the U-deletion reaction the removal of 4 U's. RNA reactants, intermediates and edited products (annotated to the right of the gels) were electrophoretically separated and densitometrically quantified. *: position of the radioactive label (^{32}P). Arrows indicate the position of the fully edited mRNA products. Fractions numbered in green were used for the SHAPE-experiments in this study

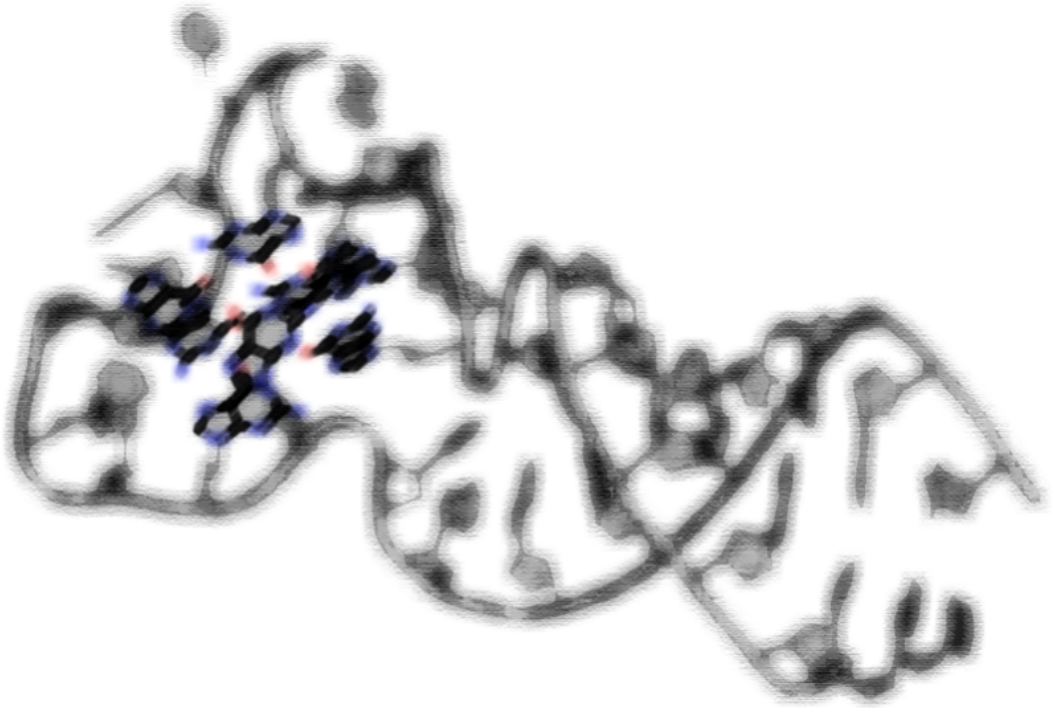


Supplementary Figure 8. "gDNA"/pre-mRNA-hybrid formation in the free and editosome-bound RNA folding states. The figure illustrates an example of the RNaseH-experiment and demonstrates all steps to derive quantitative data: (A) initial phosphorimaging output, raw densitometry traces, offset correction, background subtraction/normalization and (B) peak integration.



Chapter II

Multiple G-quartet structures in pre-edited mRNAs suggest origin of RNA editing in African trypanosomes



Abstract

Mitochondrial transcript maturation in African trypanosomes requires a U-nucleotide specific RNA editing reaction. In its most extreme form hundreds of U's are inserted into and deleted from primary transcripts to generate functional mRNAs. Unfortunately, the evolutionary origin and the biological necessity of the process have remained enigmatic. Here we report a so far unrecognized structural feature of pre-edited mRNAs, which suggests a defined evolutionary origin. We demonstrate that the cryptic pre-mRNAs contain numerous clustered G-nt, which fold into G-quadruplex (GQ) structures. We identified 27 GQ's in the different pre-mRNAs and demonstrate a positive correlation between the steady state abundance of guide (g)RNAs and the sequence position of GQ elements. We postulate that the driving force for selecting G-rich sequences lies in the formation of DNA/RNA hybrid G-quadruplex (HQ) structures between the pre-edited transcripts and the non-template strands of mitochondrial DNA. HQ's are transcription termination and replication initiation sites and thus guarantee an unperturbed replication of the mt-genome. This is essential for maintaining the life cycle of the parasite. In the transcription-on state, the identified GQ's require editing as a GQ-resolving activity suggesting that the two processes coevolved and identifying the cell- and life-cycle of the parasite as evolutionary forces.

Introduction

African trypanosomes are single cell blood-parasites and as such they are of medical importance (Kennedy, 2013). The parasites go through a life cycle that involves a mammalian host and tsetse flies as transmitting vectors. Differentiation between the two life cycle stages is accompanied by considerable metabolic changes and specifically in the tsetse fly the parasites require fully functional mitochondria to synthesize ATP. Importantly, trypanosomes have only one mitochondrion per cell (Fig. 1a). As a consequence, the precise duplication and inheritence of the mitochondrial genome is essential for the parasite to maintain its life cycle (Gluenz *et al*., 2011). The mitochondrial genome, also known as kinetoplastid (k)DNA, is organized as a macromolecular network of two types of interlocked circular DNA molecules: a heterogenous population of several thousand 1kb-size, guide (g)RNA-encoding minicircles and ≤50 identical copies of so-called maxicircles (Jensen & Englund, 2012). Each maxicircle has a molecular size of 23kbp. It carries the genetic information for two ribosomal RNAs (9S, 12S), one ribosomal protein (S12) and seventeen additional protein coding sequences (Fig. 1c). Sequence conservation of the predicted polypeptides indicates that they represent a subset of subunits of the electron transport and oxidative phosphorylation systems. Importantly, 12 of the 18 open reading frames (ORF) require RNA editing in order to be converted into translatable mRNAs (Benne *et al*., 1986; Feagin *et al*., 1988; Göringer, 2012). For nine of these ORF's this involves the site-specific insertion and deletion of literally hundreds of U-nucleotides, a phenomenon that has been dubbed pan-editing (Maslov *et al*., 1994). Pan-editing increases the U-nucleotide content of the different transcripts on average from 27% to almost 60% (Fig. 1b). While the change in the U-nt content of the pre-edited mRNAs is without precedence, it has resulted in a U-centric perspective on the phenomenon (Fig. 1d). However, alternative perceptions are possible and here we advocate for a G-nt focussed viewpoint (Fig. 1e).

Results

G-nucleotide cluster analysis of pre-edited mRNAs. Pan-edited pre-mRNAs in their unedited state lack substantial nucleotide information (on average 45%). Furthermore, they are characterized by an unusual high G-content. On average 34% of the nucleotides are G's (Fig. 1b; Supplementary Table 1). As a consequence the purin/pyrimidine (R/Y) ratio of the transcripts varies between 1.5≤R/Y≤2.7 with a mean of R/Y=2. Editing reduces the G-content to an average of 19% and a mean R/Y of 0.6. Moreover, the different pre-edited transcripts contain high numbers of tracts of G-nucleotides (2≤G≤8): 67% of all G's are arranged in clusters and editing reduces the number of G-runs to 25% (Fig. 1f). This is a unique feature of G-nucleotides in these transcripts and as such the editing reaction can be viewed as a process rendering the G-cluster propensity of the different pre-edited mRNAs (Fig. 1e,f).

Pre-edited transcripts contain multiple GQ-elements. Importantly, G-rich DNA and RNA sequences have the ability to fold into thermodynamically very stable, non-canonical, higher-order structures known as G-quadruplex (GQ)-folds (Fig. 2a,b). GQ's consist of multiple (≥2) stacked arrays of four Hoogsteen-bonded guanosine nucleotides (G-tetrads) that are stabilized by potassium cations. Potential GQ-forming sequence motifs have been identified in many species ranging from bacteria to animals and have been

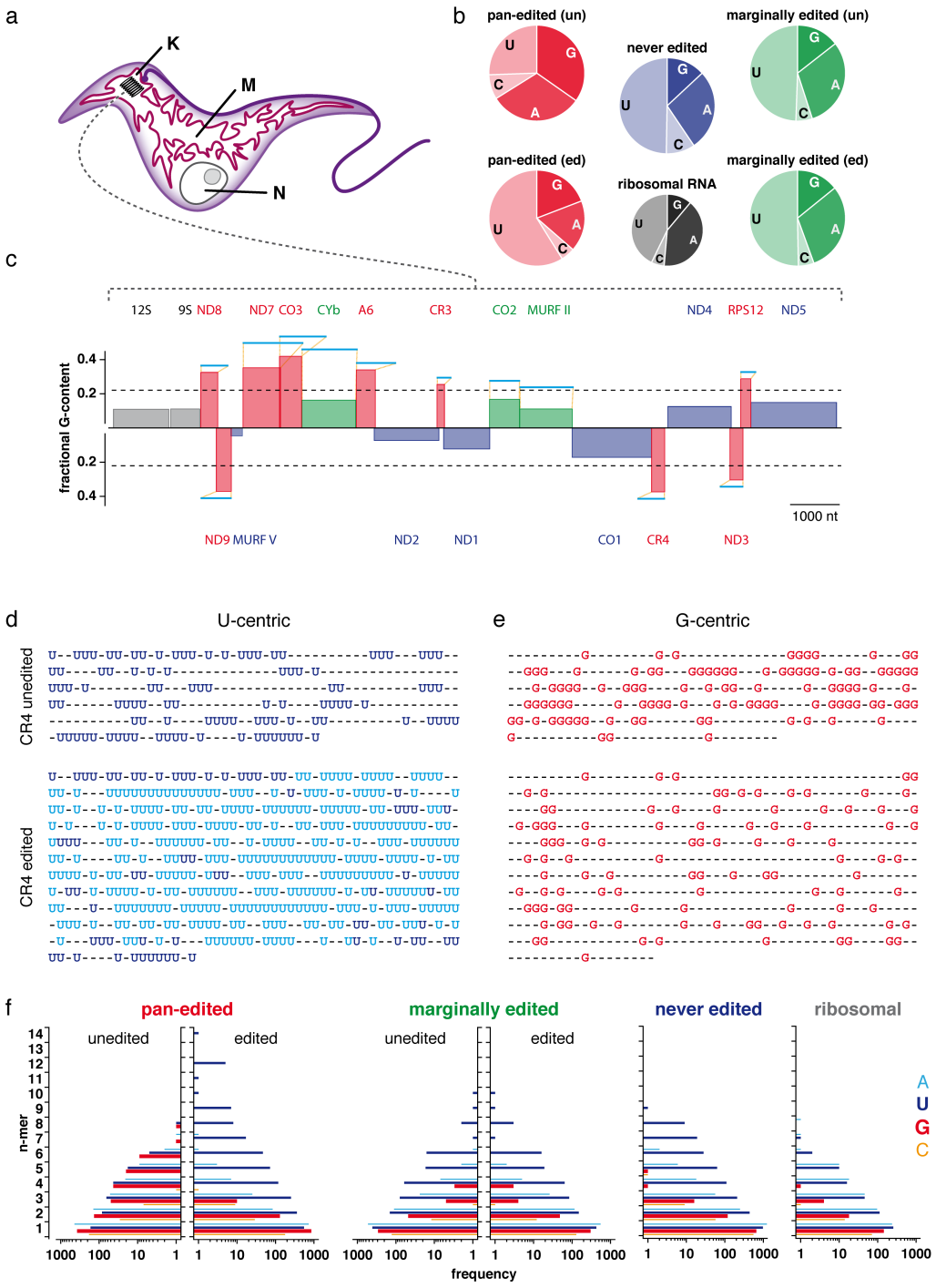


Figure 1. ***T. brucei*** **mitochondrial genome organization and nucleotide propensities.** (a) Sketch of an insect-stage *T. brucei* cell highlighting the single, extended tubular mitochondrion (M) in maroon. N: nucleus; K: kinetoplast mitochondrial DNA. (b) Nucleotide content of the different classes of mitochondrial transcripts. Red: pan-edited RNAs in their unedited (un) and edited (ed) state. Green: marginally edited RNAs in their unedited (un) and edited (ed) state. Blue: never-edited transcripts. Grey: ribosomal RNAs. (c) Linear map of both strands of the coding region of the *T. brucei* mitochondrial DNA maxicircle. Genes are annotated as boxes and are colored as in (b). Ribosomal RNA genes: 9S, 12S; pan-edited genes: ND8, ND7, CO3, A6, RPS12, ND3, CR4, ND9; marginally edited genes: Cyb, CO2, MURF II; never-edited genes: ND4, ND5, CO1, ND1, ND2, MURF V. The height of the individual boxes indicates the fractional G-content of the different ORF's. Dashed line: average G-content of 22%. All pan-edited genes are above the mean. Lines in light blue above the individual genes indicate the nt-length of the fully edited transcripts (for details see Supplementary Table 1). (d) "U-centric" view of the unedited (top) and edited (bottom) transcript of CR4. Dark blue: U-nucleotides in the pre-edited mRNA. Light blue: Inserted U's as a result of RNA editing. (e) "G-centric" view of the unedited (top) and edited (bottom) CR4-transcript emphasizing the conversion from a high number of G-runs (2≤G≤6) in the unedited state to a low number of G-tracts in the edited state. (f) Homopolymer cluster (n-mer)-analysis of all pan-edited, marginally edited and never-edited RNAs in *T. brucei* (Light blue: A-nt, dark blue: U-nt, red: G-nt, orange: C-nt). The editing-mediated reduction of homopolymer runs is unique to G-nt.

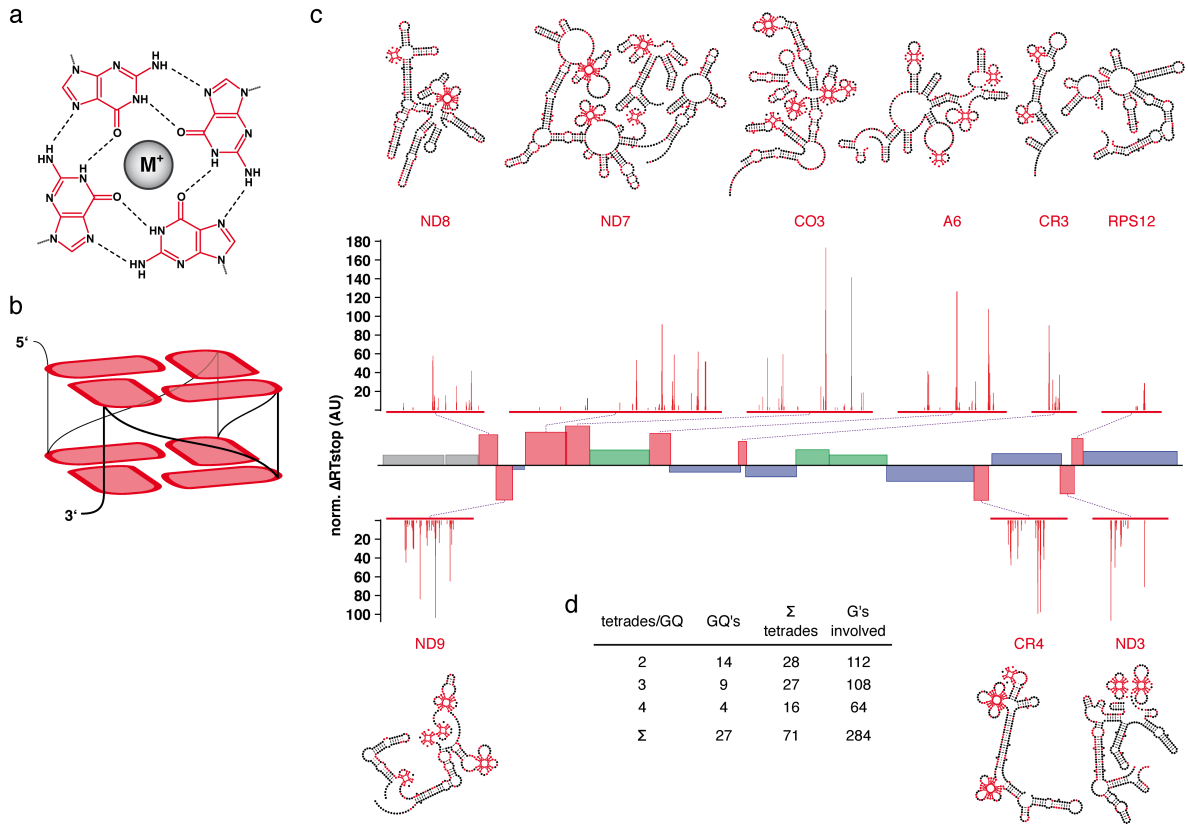


Figure 2. Experimental identification of GQ-folds. (a) Structure of a G-tetrade consisting of four Hoogsteen-bonded (black dashed lines) G-residues (red) coordinated by a stabilizing monovalent cation (M^+). (b) Parallel, intramolecular G-quadruplex formed by two stacked G-tetrades. Black line: phosphate/sugar backbone. Loop nucleotides are not shown for clarity. (c) Differential normalized reverse transcriptase (norm. ΔRT)-stop profiles (red) and predicted minimal free energy (MFE)-2D-structures of all pan-edited *T. brucei* transcripts in their pre-edited folding state. The individual structures are shown above and below the linear map of the coding region of the *T. brucei* maxicircle DNA as introduced in Fig. 1. Grey: ribosomal RNA genes; red: pan-edited genes; green: marginally edited genes; blue: never-edited genes. GQ-folds are drawn as "leaf-like" structures in red. AU: arbitrary unit. (d) Summary of the characteristics of all GQ-elements in the pan-edited transcripts.

implicated in a variety of phenomena including genome instability, telomere maintenance and the initiation of DNA-replication (Rhodes & Lipps, 2015). Based on the high number of clustered G's in the pan-edited transcripts of *T. brucei* we analyzed both strands of the maxicircle coding region for putative GQ-forming sequences using predictive algorithms. Supplementary Fig. 1b summarizes the result of a QGRS (Quadruplex-forming-G-Rich-Sequences) Mapper analysis (Kikin *et al.*, 2006) using the generic GQ-motif $G_{>2}N_{y1}G_{>2}N_{y2}G_{>2}N_{y3}G_{>2}$. Strikingly, we identified a very high number of putative GQ-forming G-nt almost exclusively in the nine pan-edited transcripts (ND8, ND9, ND7, CO3, A6, CR3, CR4, ND3 and RPS12). The same result was obtained using an energy function-based RNA-folding algorithm (ViennaRNA (Lorenz *et al.*, 2013)) and was further supported by analyzing the mitochondrial transcriptomes of the related kinetoplastid organisms *Trypanosoma cruzi* and *Leishmania tarentolae.* Although the extent of editing is lower in *L. tarentolae*, potential GQ-forming G-nt exclusively map to the pan-edited transcripts ND8, ND9, CR3, CR4, ND3 and RPS12 (Supplementary Fig. 2).

Due to the inherent possibility of false-positive and false-negative predictions we went on and identified the number of GQ-folds by experimental means. For that we used a reverse transcriptase (RT) stop assay (Hagihara *et al.*, 2010). In conjunction with DNA-sequencing the assay enables the positional mapping of GQ-folds

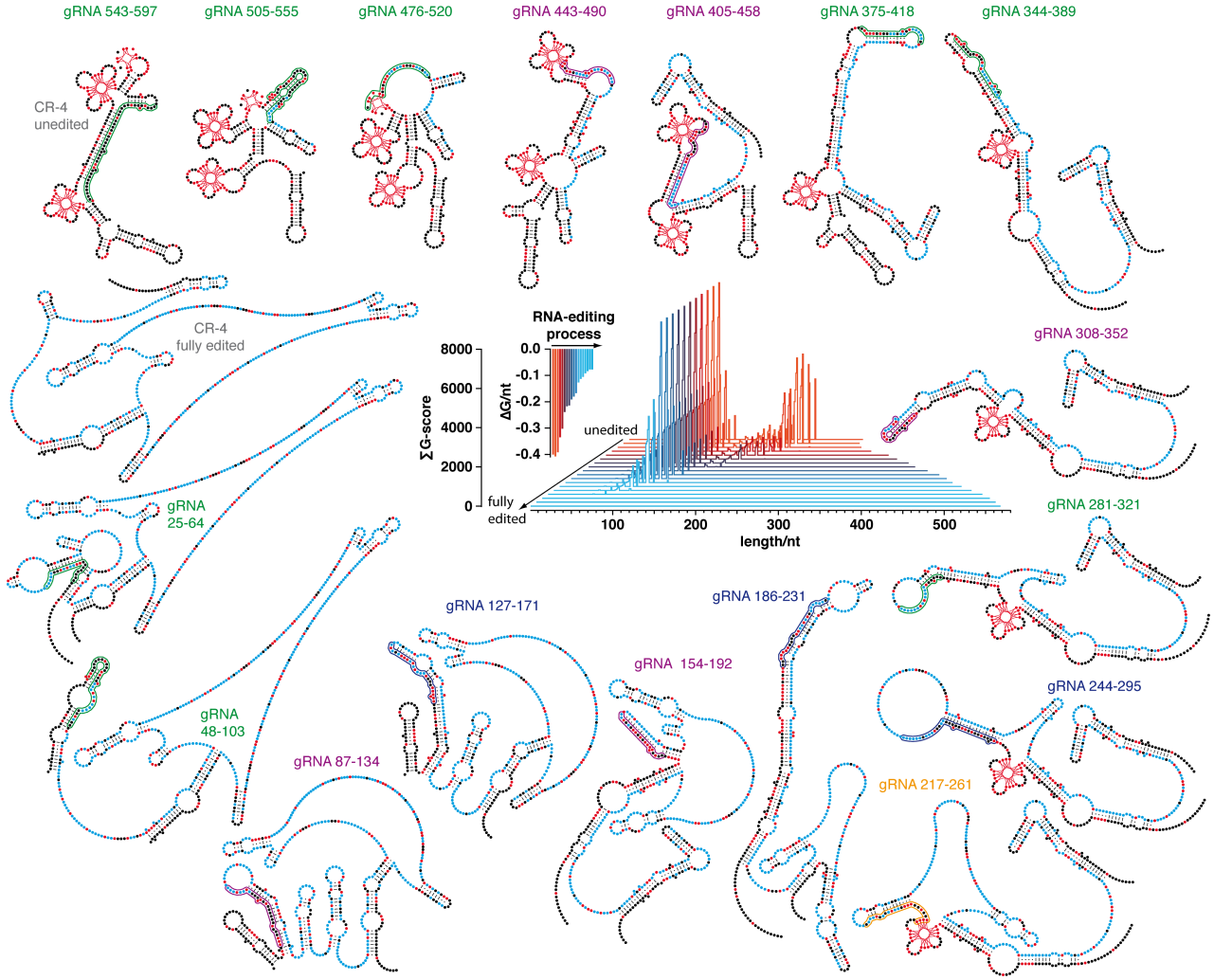


Figure 3. RNA editing as a GQ-resolving process. MFE-2D-structures of the CR4-transcript from its unedited to its fully edited folding state over 17 gRNA-driven reaction steps (clockwise starting in the upper left corner). Unedited CR4-RNA contains 3 GQ-elements ("leaf-like" structures in red), which are progressively resolved to generate a fully edited, GQ-free mRNA. The identity of the different gRNAs is given on top of the different 2D-structures. Guide RNA anchoring sites and the sequence to be edited are encircled. Colors indicate the steady state intra-mitochondrial abundances of the different gRNAs as in Koslowsky *et al.*, 2014. Inserts: ΣG-score changes and thermodynamic stability (ΔG/nt) changes of the CR4-transcript over the course of all gRNA-driven reaction steps.

with nucleotide resolution. As an additional criterion we monitored the suppression of the different RT-stop signals in the presence of Na^+- and Li^+-ions since G-quartets fold in a cation-dependent manner (K^+>>Na^+>Li^+). Fig. 2c shows a summary of the experiments. *In toto* we mapped 27 GQ-folds in the 9 pan-edited transcripts. With the exception of the RPS12 transcript (Leeder *et al.*, 2016) all pre-mRNAs contain minimally 2 GQ's with an average of 3.0 GQ's/pre-mRNA. The highest number was identified in the ND9- and CO3-transcripts, which hold 5 GQ's. The 27 GQ's involve a total of 284 G-nt and form 71 G-tetrades. This caculates to an average of 2.6 *i.e.* 3 G-tetrades/GQ-fold (Fig. 2d). Almost all RT-stop profiles are in line with the presence of ensembles of GQ-containing 2D-structures involving different G-nt within the clustered G-runs. In some cases mutually exclusive GQ's exist (Supplementary Fig. 3).

RNA editing is a GQ resolving process. As a next step we analyzed the structural fate of the different GQ's as a consequence of editing. For that we calculated the G-score- and 2D-structure-changes of

Table 1. Correlation between gRNA abundance and their ability to resolve G-tracts. On average, 79% of the G-tracts in the different pan-edited transcripts become resolved through the top 35% of gRNAs in the *T. brucei* gRNA-transcriptome (Koslowsky *et al*., 2013).

G-tracts disrupted by the 35% most abundand gRNA's

resolved G-tracts/pre-mRNA	n-mer						
	2	3	4	5	6	7	8
A6	10	4	3	1	0	0	0
CR3	3	2	0	1	0	0	0
CR4	3	2	4	3	2	0	0
CO3	17	9	7	0	1	0	0
ND3	5	4	1	0	1	0	0
ND7	15	6	6	4	1	0	1
ND8	8	1	4	3	0	1	0
ND9	7	3	3	3	2	0	0
RPS12	5	2	1	0	1	0	0
G-tracts/pre-mRNA							
A6	21	6	5	1	1	0	0
CR3	5	4	0	1	0	0	0
CR4	8	2	6	4	2	0	0
CO3	31	12	9	1	1	0	0
ND3	6	4	3	1	1	0	0
ND7	30	11	10	5	1	0	1
ND8	14	1	5	3	0	1	0
ND9	13	7	4	3	2	0	0
RPS12	8	3	1	0	1	0	0
Σ G-tracts	136	50	43	19	9	1	1
Σ resolved G-tracts	73	33	29	15	8	1	1
fraction resolved	0.54	0.66	0.67	0.79	0.89	1	1

the transcripts over the course of all gRNA-directed editing steps. Fig. 3 shows as an example the analysis of the CR4 transcript. In its pre-edited form the pre-mRNA contains 3 GQ's and requires 17 gRNAs to become fully edited. During the multistep reaction the cummulative G-score (ΣG-score) of the partially edited mRNAs progressively decreases ultimately dropping to zero in the fully edited mRNA. Similarly, the thermodynamic stabilities (expressed as ΔG/nt) decrease from -0.4kcal/mol to -0.08kcal/mol and this trend is identical for all pan-edited transcripts in *T. brucei*, *L. tarentolae* and *T. cruzi*. Thus, the gRNA-directed insertion and deletion of U-residues resolves the individual GQ-elements in a stepwise fashion to generate less structured ORF's. Eleven of the 17 gRNAs (65%) are required to resolve the 3 GQ's of pre-edited CR4, which demonstrates that the U-insertion/deletion reaction not only serves to create a translatable ORF, it also functions to eliminate GQ-folds. This is further substantiated by the observation that the steady state abundance of gRNAs correlates with the position(s) of G-tracts in the different transcripts. The top 35% of gRNAs (Koslowsky *et al*., 2014) resolve 79% of all G-tracts between 2 and 8 G's (2≤G≤8). This value is even more striking for G-runs of 4≤G≤8 (87%) or tracts of 5≤G≤8 (92%) (Table 1).

Evolutionary implications

Importantly, this so far unrecognized structural feature of all pan-edited *T. brucei* transcripts permits a new outlock onto the enigmatic evolutionary origin of the process (Speijer, 2008; Gray, 2012). If editing in part acts as a structural repair mechanism to generate GQ-free transcripts, then any potential function of GQ-folds must be located upstream of the processing reaction. GQ's in RNAs have been identified as transcription termination and replication priming sites (Wanrooij *et al*., 2010). These functions are mediated through the formation of DNA/RNA hybrid G-quartet (HQ)-structures between the nascent transcript and the non-template DNA strand (Wanrooij *et al*., 2012; Zhang *et al*., 2014; Zheng *et al*., 2014) . Thus, we hypothesize that the primary function and evolutionary advantage of the G-rich sequence motifs in the pan-edited transcripts lies in the formation of HQ-folds with the non-coding strand(s) of the maxicircle DNA. This generates multiple transcription termination sites and perhaps replication primers to assure the unperturbed replication of the mitochondrial genome. As demonstrated for human mitochondria, a separation of mitochondrial transcription and replication avoids that the two synthesizing machineries collide on the circular DNA template, thereby increasing the processivity and fidelity of both processes (Agaronyan *et al*., 2015). This is even more important in African trypanosomes for two reasons: First, the organism has only one

mitochondrion and second, a functional mitochondrion is absolutely required in the fly stage of the parasite life cycle. Cells with non-functional mitochondria or akinetoplastid parasites cannot survive in the insect as evidenced by the *T. brucei* subspecies *T. evansi* and *T. equiperdum* (Carnes *et al.*, 2015). However, the selectional advantage of HQ-elements to favour mitochondrial replication creates a structural obstacle during all cell cycle stages requiring mitochondrial transcription. In the transcription-on state the G-rich sequences ultimately adopt the above identified GQ-folds, in addition to the fact that the cryptic pre-mRNAs lack substantial sequence information. Thus, the reduction in nucleotide complexity of mtDNA to favor the frequent formation of HQ-type transcriptional stops, inevitably requires a GQ-fold resolving and RNA sequence restoration activity in the transcription-on state. As a consequence we hypothesize that the two processes co-evolved and that HQ- and GQ-folds define mutually exclusive states of mitochondrial activity: HQ-elements identify the replication-on/transcription-off state and GQ-structures classify the replication-off/transcription-on situation, which requires RNA editing (Fig. 4).

The formation of HQ-folds is likely restricted to the kinetoplast S-phase, which precedes DNA-duplication of the nuclear genome (Gluenz *et al.*, 2011; Jensen & Englund, 2012). Unfortunately, the molecular regulators that control the cell cycle timing of mitochondrial DNA-replication in *T. brucei* are only marginally understood (Jensen & Englund, 2012; Jones *et al.*, 2014). For instance, it is not known whether protein factors homologous to the TEFM-protein in human mitochondria contribute to the replication/transcription switch (Agaronyan *et al.*, 2015). However, a cell cycle-dependence of RNA editing has been suggested for *L. tarentolae* (Carrillo *et al.*, 2001) and we propose that the *T. brucei* editing accessory factor MRB1590 represents a mitochondrial GQ-interacting protein (Shaw *et al.*, 2015). Additional support comes from the observation that the G-quadruplex binding compound diminazene aceturate (Zhou *et al.*, 2014) inhibits mitochondrial DNA replication in *Trypanosoma cruzi* (Zuma *et al.*, 2015). As predicted by our model this generates dyskinetoplastid parasite cells and reinforces a large body of earlier work demonstrating that kinetoplastid DNA and perhaps G-quadruplex structures represent suitable targets for a therapeutic intervention against trypanosomatid parasites (Trager & Rudzinska, 1964; Guttman & Eisenman, 1965).

Lastly, our hypothesis is consitent with the observation that U-insertion/deletion-type RNA editing is restricted to kinetoplastid protist organisms as an early branch of the mitochondria-

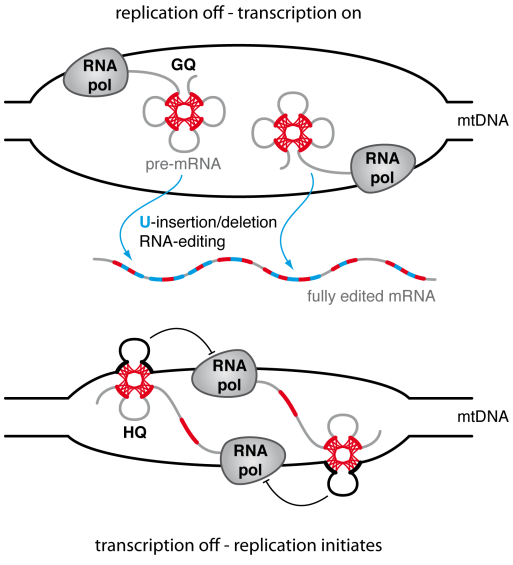


Figure 4. Switching between mitochondrial transcription and replication. Artistic rendering of the transcription-on/replication-off and transcription-off/replication-on phases of the *T. brucei* maxicircle DNA. GQ-elements in the synthesized transcripts define the transcription-on/replication-off state and require RNA editing to structurally resolve the different GQ-elements to generate translatable mRNAs (upper panel). The formation of HQ-folds between the nascent transcripts and the non-template strand(s) of the DNA maxicircle are characteristic for the transcription-off/replication-on situation (lower panel). GQ- and HQ-folds are shown as connected "leaf-like" structures. RNA pol: RNA polymerase.

containing eukaryotic lineage. While we cannot rule out some aspects of the "gene fragmentation model" of Speijer, 2008 (Speijer, 2008), our data exclude "constructive neutral evolution" as a possible scenario (Gray, 2012) and instead propose an interdependence of positive selection and coevolution as driving forces for RNA editing.

Materials and methods

Gene cloning and synthesis of pre-edited mRNAs. Mitochondrial genes encoding subunit 6 of the mitochondrial ATPase (A6), NADH dehyrogenase subunits 3, 5, 7, 8 and 9 (ND3, ND5, ND7, ND8, ND9), cytochrome C oxidase subunits 1 and 3 (CO1, CO3), C-rich regions 3 and 4 (CR3, CR4) and ribosomal protein S12 (RPS12) were PCR-amplified from *T. brucei* genomic DNA (strain Lister 427) followed by cloning into the SacI and KpnI restriction endonuclease sites of phagemid pBS SKII$^-$ (Invitrogen). Sequences were verified by DNA sequencing and RNA transcripts were generated by run-off transcription using T7 RNA polymerase and linearized plasmids following standard protocols. Samples were DNaseI treated (37°C, 15min) followed by phenol extraction. Non-incorporated rNTP's were removed by size exclusion chromatography and RNA transcripts were recovered by EtOH precipitation. RNAs were dissolved in 10mM Tris/HCl pH 7.5, 1mM EDTA (TE) and RNA concentrations were calculated from UV-light absorbancy measurements at 260nm. The homogeneity and molecular sizes of the different RNAs was analyzed electrophoretically in 6% (w/v) urea-containing (8M) polyacrylamide gels (Supplementary Fig. 1a).

Identification of potential GQ-forming sequences. *In silico* predictions of GQ-structures were performed for all *T. brucei* and *L. tarentolae* mitochondrial transcripts in their edited and unedited states. RNA sequences were retrieved from the U-insertion/deletion Edited-Sequence-Database (http://dna.kdna.ucla.edu/trypanosome/files/tbma-xi.html) and analyzed using QGRS Mapper (Kikin *et al*., 2006) and ViennaRNA 2.2.3 (Lorenz *et al*., 2013). QGRS Mapper predictions were performed using the low stringency constraints of ≥2 stacked G-tetrads and loop sizes between 0-36nt. The likelihood of forming stable GQ-elements was quantified by calculating "G-score"-values (Kikin *et al*., 2006), which were summed up to yield cumulative (Σ)G-scores per nt position. ViennaRNA-based predictions relied on a thermodynamic folding algorithm using the energy function: E [L,l] = a(L−1) + b x ln(l−2). L: number of G-tetrades (2≤L≤5); l: connecting loop length; a(37°C): unpaired nucleotide penalty of −18kcal/mol; b(37°C): loop strain parameter of 12kcal/mol. MFE-secondary structures and thermodynamic stabilities were extrapolated to 27°C, which is the optimal growth temperature of editing-active, insect-stage trypanosomes.

Experimental verification of GQ-folds and data analysis. The ability of G-clusters to adopt GQ-folds was experimentally verified at single nucleotide resolution using a differential reverse transcriptase (RT) stop assay at GQ-favouring (K^+) and GQ-disfavouring (Na^+, Li^+) cation conditions. RT-stop reactions were performed using 1pmol pre-edited transcripts and equimolar amounts of 5'-fluorescenctly labeled DNA-oligonucleotide primer molecules. After heat denaturation (95°C, 1.5-2.5min, 0.25xTE pH7.5) DNA-primers were annealed at 50-65°C (7.5min) followed by snap-cooling. Reverse transcription was performed by adding a concentrated reaction mix to yield final concentrations of 50nM transcript, 50nM DNA primers, 50mM Tris/HCl pH8.3, 3mM $MgCl_2$, 5mM

dithiothreitol, 0.75U/µL RiboLock™ and 0.5mM dNTP (each). Monovalent cation concentrations were variied between 1mM and 75mM until a sufficient readthrough was achieved. Some experiments required 0.75mM Ni^{2+} as GQ-destabilizing cation (Blume *et al*., 1997). Reaction mixes were incubated at 52°C for 15-20min at a SuperScriptIII™ RT (ThermoFisher Scientific) concentration of 10U/µL. RT-reactions were terminated by snap cooling followed by RNA hydrolysis in 400mM NaOH (95°C, 5min). Resulting cDNAs were EtOH-precipated, resuspended in Hi-Di™ formamide and analyzed by capillary electrophoresis (CE). Raw CE-traces were baseline adjusted and signal decay corrected using ShapeFinder (Vasa *et al*., 2008). Peak identities were determinded by cDNA-sequencing followed by Gaussian peak integration. Pronounced stops were identified by boxplotting as peak integrals ≥1.25-fold the interquartile range. Remaining values were averaged and all peak integrals were divided by the average to normalize small peaks to ~1. GQ-induced stops were defined as stops showing a differential (Δ) response ≥2.5-fold the average peak integral in GQ-favouring (K^+) *versus* GQ-disfavouring (Na^+, Li^+) conditions. Per transcript minimally 5 independent experiments were performed (mean r=0.83). For a detailed analysis see Supplementary Fig. 4.

RT-stop guided RNA 2D-structure prediction. RNA 2D-structures were generated with the help of ViennaRNA 2.2.3 (Lorenz *et al*., 2013) using the identified RT-stop positions to predict the folding and thermodynamic stability of the different GQ-elements. In some cases the effective loop length (l) was reduced by allowing long loops to adopt additional 2D-structure. GQ-sequence stretches with the lowest Gibbs free energies were forced into a single-stranded conformation during the subsequent 2D-structure prediction and were manually introduced to create RT-stop data refined minimal free energy (MFE) models. To estimate the Gibbs free energy of a given GQ-containing RNA structure all individual energy terms were summed up.

References

Agaronyan K, Morozov YI, Anikin M, Temiakov D. 2015. Mitochondrial biology. Replication-transcription switch in human mitochondria. Science. 347:548-551.

Benne R, Van den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC. 1986. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell. 46:819-826.

Blume SW, Guarcello V, Zacharias W, Miller DM. 1997. Divalent transition metal cations counteract potassium-induced quadruplex assembly of oligo(dG) sequences. Nucleic Acids Res. 25:617-625.

Carnes J, Anupama A, Balmer O, Jackson A, Lewis M, Brown R, Cestari I, Desquesnes M, Gendrin C, Hertz-Fowler C, Imamura H, Ivens A, Kořený L, Lai DH, MacLeod A, McDermott SM, Merritt C, Monnerat S, Moon W, Myler P, Phan I, Ramasamy G, Sivam D, Lun ZR, Lukeš J, Stuart K, Schnaufer A. 2015. Genome and phylogenetic analyses of *Trypanosoma evansi* reveal extensive similarity to *T. brucei* and multiple independent origins for dyskinetoplasty. PLoS Negl Trop Dis. 9:e3404.

Carrillo R, Thiemann OH, Alfonzo JD, Simpson L. 2001. Uridine insertion/deletion RNA editing in *Leishmania tarentolae* mitochondria shows cell cycle dependence. Mol Biochem Parasitol. 113:175-181.

Feagin JE, Abraham JM, Stuart K. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell. 53:413-422.

Gluenz E, Povelones ML, Englund PT, Gull K. 2011. The kinetoplast duplication cycle in *Trypanosoma brucei* is orchestrated by cytoskeleton-mediated cell morphogenesis. Mol Cell Biol. 31:1012-1021.

Göringer HU. 2012. 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. Annu Rev Microbiol. 66:65-82.

Gray MW. 2012. Evolutionary origin of RNA editing. Biochemistry. 51:5235-5242.

Guttman HN, Eisenman RN. 1965. Acriflavin-induced loss of kinetoplast deoxyribonucleic acid in *Crithidia fasciculata* (Culex pipiens strain). Nature. 207:1280-1281.

Hagihara M, Yoneda K, Yabuuchi H, Okuno Y, Nakatani K. 2010. A reverse transcriptase stop assay revealed diverse quadruplex formations in UTRs in mRNA. Bioorg Med Chem Lett. 20:2350-2353.

Jensen RE, Englund PT. 2012. Network news: the replication of kinetoplast DNA. Annu Rev Microbiol. 66:473-491.

Jones NG, Thomas EB, Brown E, Dickens NJ, Hammarton TC, Mottram JC. 2014. Regulators of *Trypanosoma brucei* cell cycle progression and differentiation identified using a kinome-wide RNAi screen. PLoS Pathog. 10:e1003886.

Kennedy PG. 2013. Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). Lancet Neurol. 12:186-94.

Kikin O, D'Antonio L, Bagga PS. 2006. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. 34(Web Server issue):W676-82.

Koslowsky D, Sun Y, Hindenach J, Theisen T, Lucas J. 2014. The insect-phase gRNA transcriptome in *Trypanosoma brucei*. Nucleic Acids Res. 42:1873-1886

Leeder WM, Voigt C, Brecht M, Göringer HU. 2016. The RNA chaperone activity of the *Trypanosoma brucei* editosome raises the dynamic of bound pre-mRNAs. Sci Rep. 6:19309

Lorenz R, Bernhart SH, Qin J, Höner zu Siederdissen C, Tanzer A, Amman F, Hofacker IL, Stadler PF. 2013. 2D meets 4G: G-quadruplexes in RNA secondary structure prediction. IEEE/ACM Trans Comput Biol Bioinform. 10:832-844.

Maslov DA, Avila HA, Lake JA, Simpson L. 1994. Evolution of RNA editing in kinetoplastid protozoa. Nature. 368:345-348.

Rhodes D, Lipps HJ. 2015. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 43:8627-8637.

Shaw PL, McAdams NM, Hast MA, Ammerman ML, Read LK, Schumacher MA. 2015. Structures of the *T. brucei* kRNA editing factor MRB1590 reveal unique RNA-binding pore motif contained within an ABC-ATPase fold. Nucleic Acids Res. 43:7096-7109.

Speijer D. 2008. Evolutionary aspects of RNA editing. In: RNA Editing. Springer Publishing Heidelberg. Ed. Göringer, H. U. pp. 199-227.

Trager W, Rudzinska MA. 1964. The riboflavin requirement and the effects of acriflavin on the fine structure of the kinetoplast of *Leishmania tarentolae*. J Protozool. 11:133-145.

Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. 2008. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA. 14:1979-1990.

Wanrooij PH, Uhler JP, Simonsson T, Falkenberg M, Gustafsson CM. 2010. G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. Proc Natl Acad Sci U S A. 107:16072-16077.

Wanrooij PH, Uhler JP, Shi Y, Westerlund F, Falkenberg M, Gustafsson CM. 2012. A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. Nucleic Acids Res. 40:10334-10344.

Zhang JY, Zheng KW, Xiao S, Hao YH, Tan Z. 2014. Mechanism and manipulation of DNA:RNA hybrid G-quadruplex formation in transcription of G-rich DNA. J Am Chem Soc. 136:1381-1390.

Zheng KW, Wu RY, He YD, Xiao S, Zhang JY, Liu JQ, Hao YH, Tan Z. 2014. A competitive formation of DNA:RNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site. Nucleic Acids Res. 42:10832-10844.

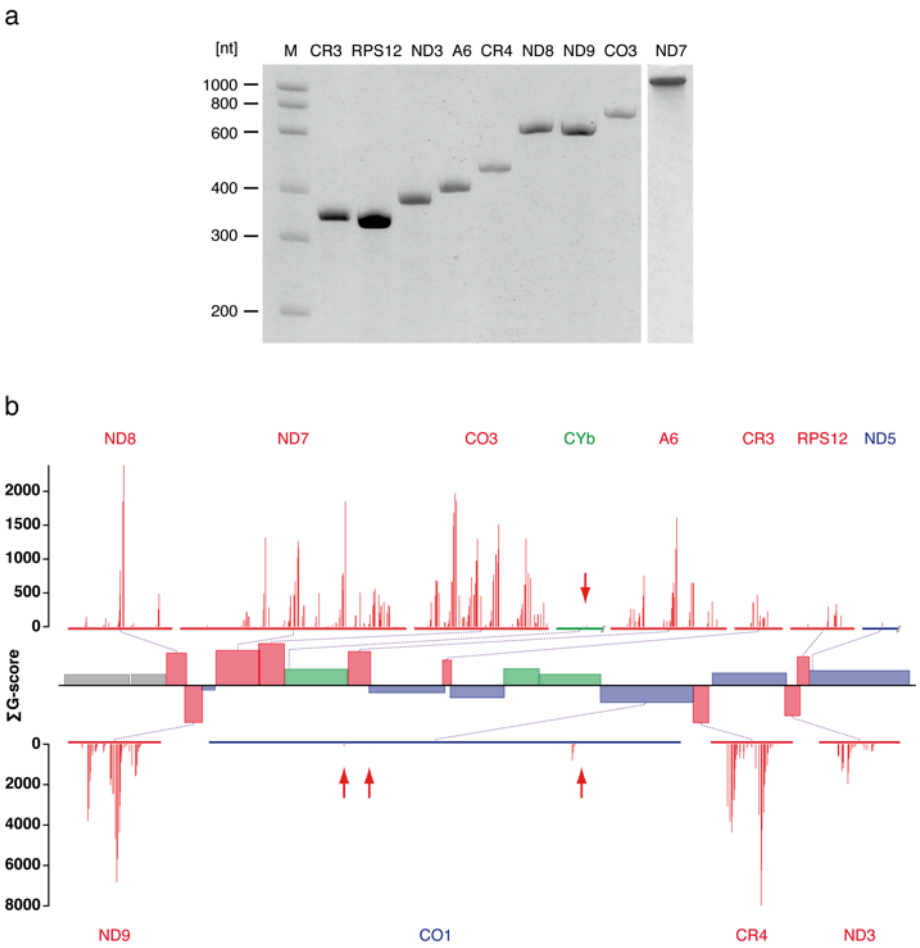
Zhou J, Le V, Kalia D, Nakayama S, Mikek C, Lewis EA, Sintim HO. 2014. Diminazene or berenil, a classic duplex minor groove binder, binds to G-quadruplexes with low nanomolar dissociation constants and the amidine groups are also critical for G-quadruplex binding. Mol Biosyst. 10:2724-34.

Zuma AA, Cavalcanti DP, Zogovich M, Machado AC, Mendes IC, Thiry M, Galina A, de Souza W, Machado CR, Motta MC. 2015. Unveiling the effects of berenil, a DNA-binding drug, on *Trypanosoma cruzi*: implications for kDNA ultrastructure and replication. Parasitol Res. 114:419-430.

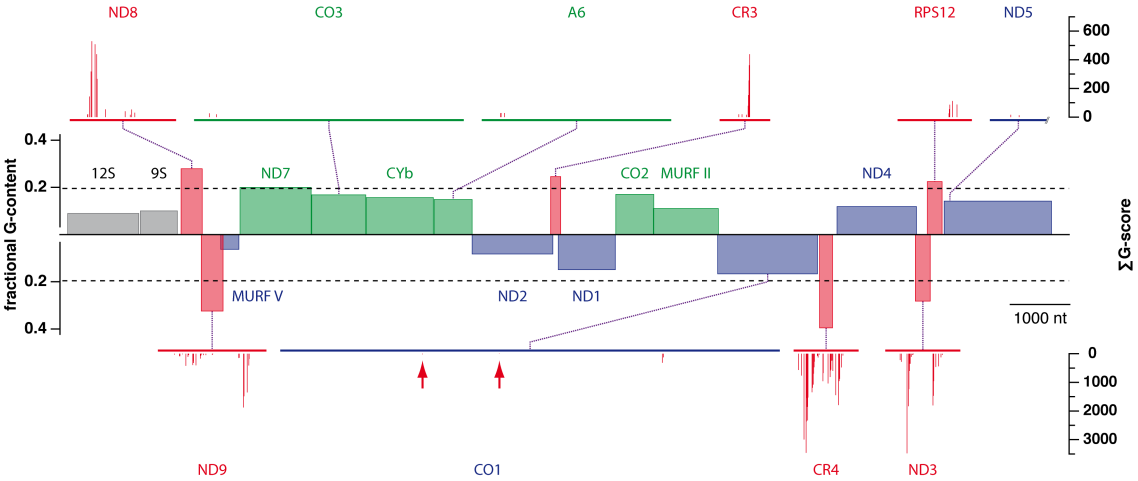
Author Contributions:
Conceived and designed experiments: HUG, WML
Performed experiments: WML, NFCH

Supplementary Tables/Figures

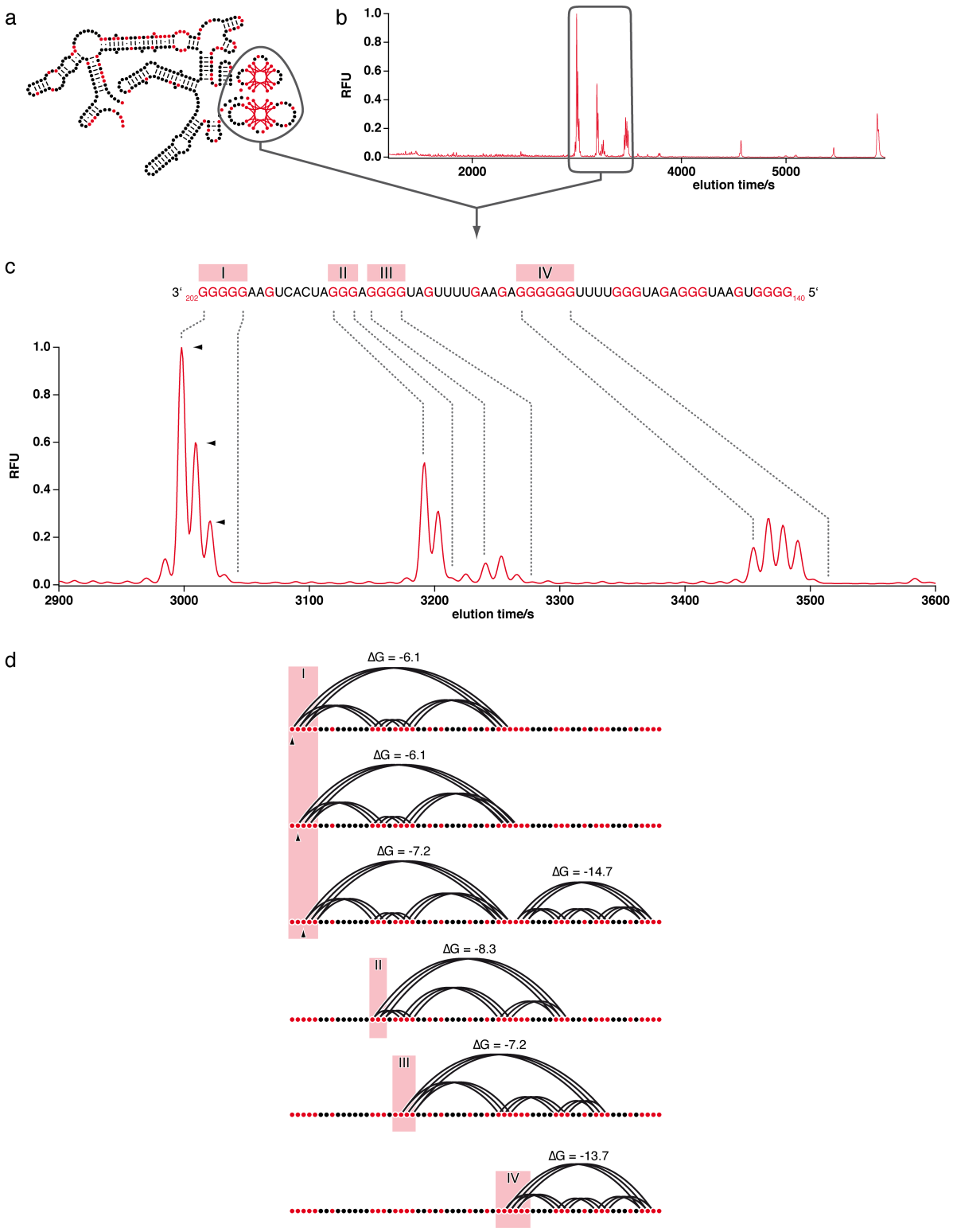
Supplementary Figure 1. RNA characterization and bioinformatic analysis of potential GQ-forming nucleotides. (a) Gelelectrophoretic characterization of the *T. brucei* CR3-, RPS12-, ND3-, A6-, CR4-, ND8-, ND9-, CO3- and ND7-transcripts used in this study. Electrophoresis was performed in 8M urea-containing 6% (w/v) polyacrylamide gels. M: molecular size marker. (b) QGRS Mapper analysis (Kikin *et al.*, 2006) of potential GQ-forming G-nt in the coding regions of the *T. brucei* mitochondrial genome. The figure shows a linear map of both strands of the *T. brucei* DNA maxicircle. Grey: ribosomal RNA genes (9S, 12S); red: pan-edited genes (ND8, ND7, CO3, A6, RPS12, ND3, CR4, ND9); green: marginally edited genes (CYb, CO2, MURF II); blue: never-edited genes (ND4, ND5, CO1, ND1, ND2, MURF V). The height of the individual boxes indicates the relative G-content of the different ORF's (see Supplementary Table 1). The likelihood of forming GQ-motifs was caculated using the heuristic QGRS Mapper-based "G-score"-metric (Kikin *et al.*, 2006). G-scores are non-negative intergers and cumulative (Σ)G-score profiles are plotted above and below all relevant ORF's. Positions with low ΣG-scores are marked with arrows in red.



Supplementary Figure 2. Identification of potential GQ-forming nucleotides in the mitochondrial genome of *Leishmania tarentolae*. QGRS Mapper analysis (Kikin *et al*., 2006) of potential GQ-forming G-nt in the coding region of the *L. tarentolae* DNA maxicircle. The figure shows a linear map of both strands of the maxicircle DNA. Grey: ribosomal RNA genes (9S, 12S); red: pan-edited genes (ND8, CR3, RPS12, ND3, CR4, ND9); green: marginally edited genes (ND7, CO3, CYb, A6, CO2, MURF II); blue: never edited genes (ND4, ND5, CO1, ND1, ND2, MURF V). The height of the individual boxes indicates the fractional G-content of the different ORF's (y-axis on the left). Dashed horizontal line in black: average G-content of all transcripts (20%). Pan-edited genes are above the mean. The likelihood of forming a GQ-motif was caculated using the heuristic QGRS Mapper-based "G-score" metric (Kikin *et al*., 2006). Cumulative (Σ)G-score profiles are plotted above and below all relevant ORF's (y-axis on the right). Positions with low ΣG-scores are marked with arrows in red.

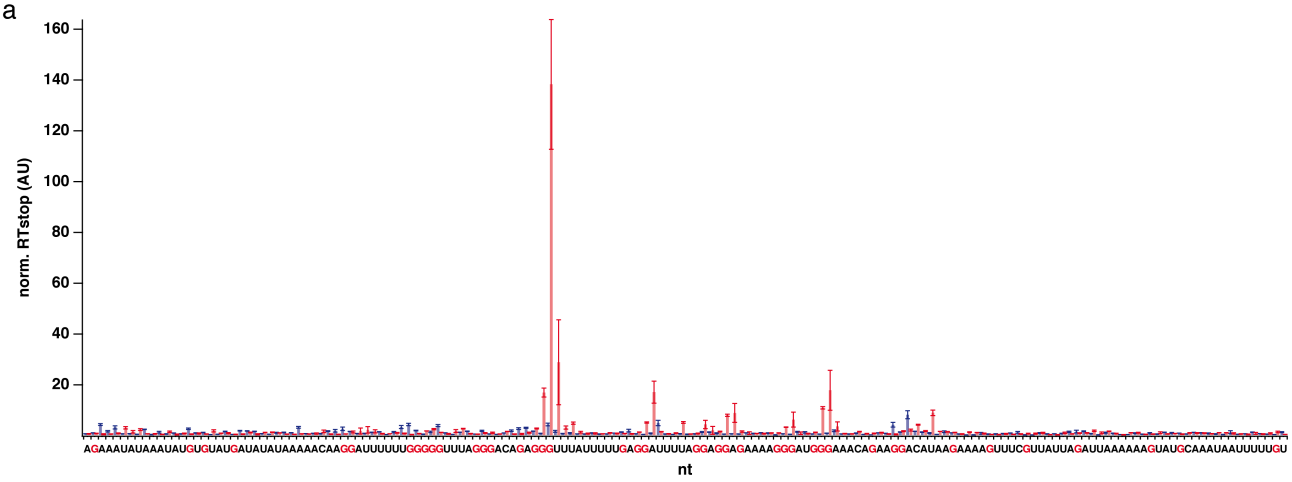


Supplementary Figure 3. RT-stop data map mutiple GQ-folds within a thermodynamic ensemble of RNA 2D-structures. (a) MFE-2D-structure of the pre-edited ND3 mRNA displaying two GQ-elements each containing three G-tetrades. (b) Corresponding RT-stop capillary electrophoresis (CE) profile at 75mM K^+. RFU: relative fluorescence unit. (c) The different peaks map to a sequence stretch between G_{140} and G_{202} containing four G-tracts (I to IV). (d) Arc representation of different GQ-elements that correlate with individual RT-stop signals (arrow heads in black) demonstrating the co-existence of GQ's of similar Gibbs free energies within an ensemble of RNA structures and in some cases of mutually exclusive GQ's. Red dots: G-nucleotides. ΔG's are in kcal/mol.

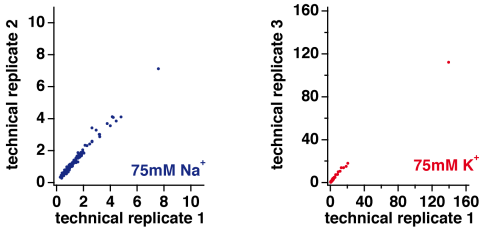


Supplementary Figure 4. Experimental variation of RT-stop experiments. (a) Normalized RT-stop data of the pre-edited CR3-transcript plotted as a function of the CR3-primary sequence. G-nt are highlighted in red. Error bars are SD. (b) Representative Pearson plots of three technical replicates of normalized RT-stop data. Blue: GQ-disfavoring (75mM Na^+) and red: GQ-favoring (75mM K^+) cation conditions. (c) Pearson (*r*) and Spearman (ρ) correlation coefficients and average SD of all RT-stop data at GQ-disfavoring and GQ-favoring cation conditions. AU: arbitrary unit. nt: nucleotide.

a



b



c

	GQ-disfavouring conditions	GQ-favouring conditions
Pearson (*r*)	0.97 ± 0.02	0.99 ± 0.01
Spearman (ρ)	0.94 ± 0.04	0.88 ± 0.03
average SD	0.11 ± 0.07	0.18 ± 0.13

Supplementary Table 1. Nucleotide propensities of all *T. brucei* mitochondrial transcripts. (a) pan-edited, (b) marginally edited and (c) never edited mRNAs. un: unedited; ed: edited. Pan-edited pre-mRNAs in their unedited state are purine-rich with R/Y ratios of 1.5≤R/Y≤2.7. Red background: above average. Blue background: below average. (d) G-nucleotides are arranged in clusters. Pan-edited mRNAs contain roughly 2.5-times more G-tracts (n≥2) when compared to transcripts that are marginally or never edited. On average these clusters are only 10nt apart, which is 4-7-fold closer than in any other RNA analyzed.

a

pan-edited

mRNA	$A6_{un}$	$A6_{ed}$	$CR3_{un}$	$CR3_{ed}$	$CR4_{un}$	$CR4_{ed}$	$CO3_{un}$	$CO3_{ed}$
length	401	820	164	300	283	568	463	970
#A	149	149	65	65	65	65	144	144
#G	137	137	42	42	107	107	195	195
#C	21	21	6	6	6	6	28	28
#U	94	513	51	187	105	390	96	603
R/Y	2.5	0.5	1.9	0.6	1.5	0.4	2.7	0.5
G's (%)	34	17	26	14	38	19	42	20

mRNA	$ND3_{un}$	$ND3_{ed}$	$ND7_{un}$	$ND7_{ed}$	$ND8_{un}$	$ND8_{ed}$	$ND9_{un}$	$ND9_{ed}$
length	268	465	783	1246	361	574	322	647
#A	87	87	244	244	100	100	99	99
#G	82	82	277	277	118	118	120	120
#C	20	20	90	90	44	44	25	25
#U	79	276	172	635	99	312	78	403
R/Y	1.7	0.6	2.0	0.7	1.5	0.6	2.1	0.5
G's (%)	31	18	35	22	33	21	37	19

b

marginally edited

mRNA	$CO2_{un}$	$CO2_{ed}$	CYb_{un}	CYb_{ed}	$MURF2_{un}$	$MURF2_{ed}$
length	632	636	1118	1152	1091	1112
#A	210	210	344	344	318	318
#G	107	107	184	184	123	123
#C	46	46	69	69	40	40
#U	269	273	521	555	610	631
R/Y	1.0	1.0	0.9	0.8	0.7	0.7
G's (%)	17	17	16	16	11	11

c

never edited

mRNA	CO1	ND1	ND4	ND5	MURF1	MURF5
length	1650	960	1314	1773	1343	234
#A	373	242	442	472	384	94
#G	286	120	167	267	101	11
#C	202	127	96	99	148	27
#U	789	471	609	935	710	102
R/Y	0.7	0.6	0.9	0.7	0.6	0.8
G's(%)	17	13	13	15	8	5

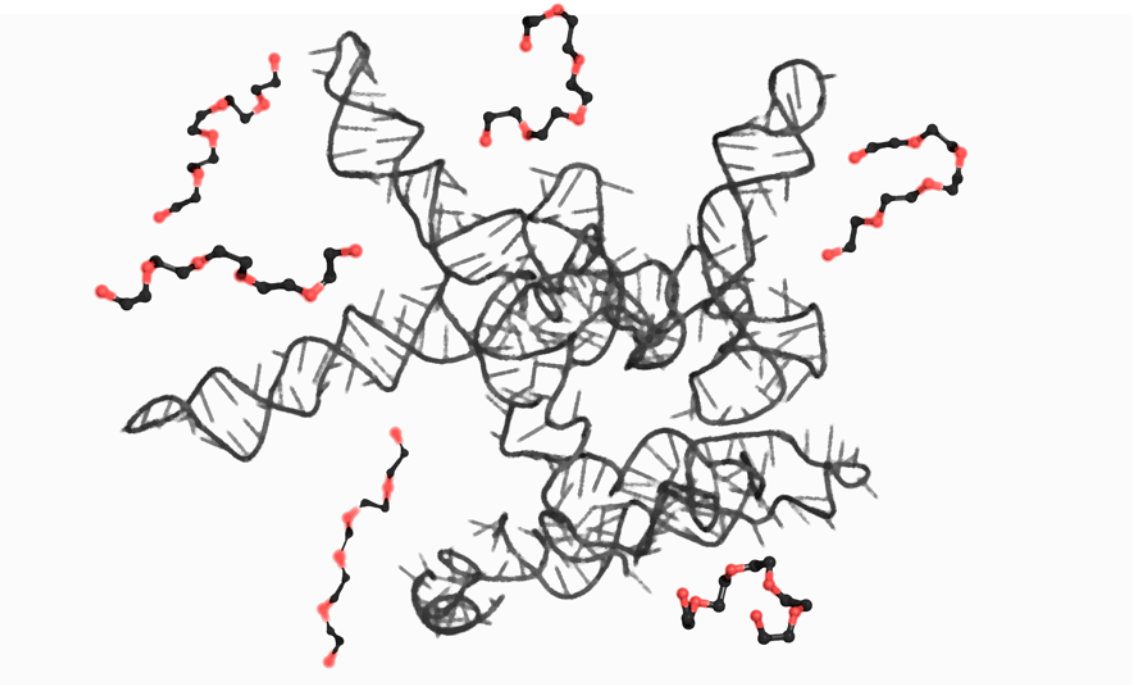
rRNA	9S	12S
length	611	1149
#A	249	462
#G	69	128
#C	35	67
#U	258	492
R/Y	1.1	1.1
G's(%)	11	11

d

mRNA	G's in G-tracts (%)	average distance between G-tracts (nt)
$pan\text{-}edited_{un}$	67.0	10
$pan\text{-}edited_{ed}$	25.4	40
$marginally\ edited_{un}$	29.7	49
$marginally\ edited_{ed}$	28.4	51
never edited	30.8	51
ribosomal RNAs	26.4	74

Chapter III

The 2D-structure of the mitochondrial RPS12 pre-mRNA is not affected by volume exclusion conditions



Abstract

Mitochondrial transcript maturation in African trypanosomes requires an RNA editing reaction in which non-functional pre-mRNAs are converted into translatable mRNAs. While the reaction is characterized by the site-specific insertion and deletion of exclusively U-nucleotides, some transcripts are edited through the insertion of hundreds of U's, showcasing the cryptic nature of the various pre-mRNAs. In addition to the lack of nucleotide information, the different RNAs are further characterized by an unusual high G-content and as a consequence the formation of thermodynamically highly stable 2D-structures in the majority of cases even involving multiple G-quadruplex (GQ)-folds. Unfortunately, it is not clear whether these structures resemble the *in vivo* folds of the different RNAs given the extreme volume exclusion *i.e.* "crowding" conditions within the trypanosome mitochondrion. Here we analyze the effect(s) of volume exclusion on the structure of the mitochondrial RPS12 pre-mRNA. We use polyethylene glycol $(PEG)_{4000}$ as a neutral macromolecular cosolute to mimick intra-mitochondrial solvent conditions and we monitor the structure of the RNA on a global scale and with nucleotide resolution. We demonstrate that macromolecular crowding has no impact on the 2D-fold of the RPS12 pre-mRNA and we conclude that the determined minimal free energy (MFE)-structure in dilute solvent conditions represent a good proxy for the folding of the RPS12 pre-mRNA within its mitochondrial solvent context.

Introduction

The folding of RNA molecules into compact, native structures or ensembles of structures is dictated by a set of first principle physicochemical forces (Doudna & Cech, 2002). One of which is charge-compensation to overcome the electrostatic repulsion of the negatively charged phosphodiester backbone (Draper *et al*., 2005; Soto *et al*., 2007; Leipply & Draper, 2010). Mono and divalent metal-ions at low to high millimolar concentrations contribute to this task (Tyrrell *et al*., 2013) and the effects of metal ion-radius and charge density have been studied in detail (Manning, 1978; Misra & Draper, 2000; Lambert *et al*., 2009). Next to metal-ions, metabolites, polyamines and osmolytes have been shown to modulate RNA structure (Miyamoto *et al*., 1993; Lambert & Draper, 2007; Bennet *et al*., 2009; Lambert *et al*., 2010; Trachman III & Draper, 2013) as well as high concentrations of macromolecules, which can occupy up to 30% of the total volume of a cellular compartment. This generates so-called "crowded" solvent conditions (Zimmerman & Trach, 1991; Ellis, 2001; Thirumalai *et al*., 2003), which in general stabilize RNA 2D- and 3D-structures through the excluded volume effect and entropy perturbation of the folding landscape (Kilburn *et al*., 2010; Kilburn *et al*., 2013; Tyrrell *et al*., 2013).

RNA editing describes a post-transcriptional modification reaction of mitochondrial pre-mRNAs that is characterized by the site-specific insertion and deletion of exclusively U-nucleotides (reviewed in Göringer, 2012). The reaction takes place within the single mitochondrion of trypanosomes, which represents the most crowded intracellular environment of eukaryotic cells. Intra-mitochondrial macromolecular concentrations reach up to 560g/L (Srere, 1980; Harve *et al*., 2010). Editing is catalyzed by a macromolecular protein complex, the 20S editosome (Göringer, 2012), which interacts with 18 mitochondrial pre-mRNAs as substrates in the processing reaction. The different transcripts encode subunits of the mitochondrial electron transport and oxidative phosphorylation chains and have been characterized by several unusual features: First, the majority of pre-mRNAs lacks substantial sequence information (on average 45%), hence they require RNA editing to be converted into translatable mRNAs. Second, the different pre-mRNAs are typified by an extraordinarily high G-content (34%), which in two thirds of the cases are clustered in tracts of G-nucleotides ($2 \leq G \leq 8$). Third, *in vitro* chemical probing studies revealed that the different pre-mRNAs adopt extraordinarily stable 2D-structures approaching the stability of structural RNAs (Leeder *et al*., 2016). Next to canonical base-pairing they contain pseudoknots and in many cases multiple G-quadruplex (GQ)-folds. However, it is not clear whether these 2D-structures resemble the *in vivo* folds given the extreme crowding conditions in the trypanosome mitochondrion. This is especially important since RNA editing *in vitro* has been shown to be sensitive to crowded solvent properties (Katari *et al*., 2013) and the fact that G-quadruplex structures are stabilized by molecular crowding (Miyoshi *et al*., 2002; Fujimoto *et al*., 2011). As a consequence, here we analyze the effect(s) of macromolecular crowding on the structure of the mitochondrial RPS12 pre-mRNA as an archetypical example of mitochondrial transcripts in African trypanosomes. We use high molecular mass polyethylene glycol (PEG) as a neutral macromolecular cosolute to mimick intra-mitochondrial solvent conditions and we monitor the structure of the RPS12 transcript by temperature-dependent UV-spectroscopy and by selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE).

Results and discussion

To study the impact of macromolecular crowding on the structure of mitochondrial pre-mRNAs in *Trypanosoma brucei* we used the primary transcript of RPS12 as a representative model RNA. The pre-mRNA molecule is 325nt long. As a pan-edited transcript it is edited throughout its entire primary sequence with 132 U-nt inserted and 28 U's deleted. The RNA has a G-nt-content of 27% and an R/Y-ratio of 1.3. Its experimentally determined minimal free energy (MFE)-2D-structure calculates to a Gibbs free energy (ΔG) of −152kcal/mol with a base-paired *versus* single-stranded nucleotide ratio

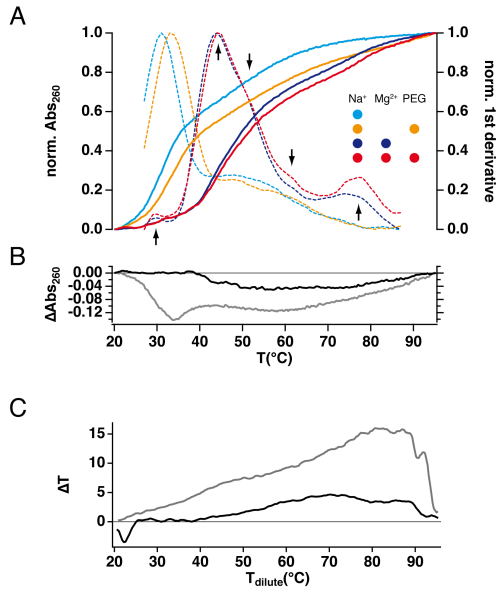


Figure 1. A: Normalized thermal UV-denaturation profiles of the RPS12 pre-mRNA from *Trypanosoma brucei* (coloured, solid lines) at different buffered solvent conditions. Red/orange: crowded solvent conditions (6% (w/w) PEG_{4000}). Blue/cyan: dilute solvent conditions. Dark colours: 2mM Mg^{2+}. Bright colours: no Mg^{2+}. Dashed traces: Normalized 1st-derivatives ($\partial A_{260}/\partial T$) of the different melting profiles. Five helix-coil transitions are marked with arrows B: Difference (Δ) UV-melting profiles (dilute minus crowded conditions) in the presence/absence of Mg^{2+}-cations (black/grey). C: Shifting of the UV-denaturation profiles to higher temperatures expressed as ΔT *vs*. T of the dilute condition in presence/absence of Mg^{2+}-cations (black/grey).

($r_{bp/ss}$) of 0.62. In addition, the RNA contains a pseudoknot (Leeder *et al*., 2016). Since *in cell* structure probing experiments have successfully only been performed with abundant, cytosolic RNAs (Spitale *et al*., 2013; Ding *et al*., 2014; Rouskin *et al*., 2014; Tyrrell *et al*., 2013; McGinnis *et al*., 2015) we decided to mimk the crowded, intra-mitochondrial solvent environment by using a chemically inert, synthetic cosolute such as polyethylene glycol (PEG) (reviewed in Minton, 2001; Chebotareva *et al*., 2004). We used PEG with a mean molecular mass of 4000g/mol (PEG_{4000}). The synthetic compound is characterized by a polymer crossover concentration (ϕ^*) of 4% (w/w), which marks the transition from a semidilute to a crowded solvent regime (De Gennes, 1979; Kozer & Schreiber, 2004; Kozer *et al*., 2007). As a consequence, all experiments were performed at 6% (w/w) PEG_{4000}.

As a first comparison, we measured UV-melting profiles of the RPS12 pre-mRNA in dilute and crowded solvent conditions. Representative normalized melting profiles and their 1st-derivatives ($\partial A_{260}/\partial T$) are shown in Fig. 1A. At dilute buffer conditions (in the presence of 70mM Na^+ and 2mM Mg^{2+}), the pre-mRNA displays a complex melting profile with 5 distinct helix/coil transitions: Two dominant transitions with melting midpoints (T_m-values) at 46°C and 78°C and three minor transitions at 30°C, 53°C and 62°C. Almost identical traces were recorded at crowded solvent conditions (Fig 1A). As demonstrated in the difference melting-curve in Fig. 1B, the two profiles superimpose perfectly at temperatures ≤40°C and deviate only slightly above 40°C. At that temperature the "crowded" profile shifts to higher temperature values, however only marginally with a ΔT of maximally 4°C (Fig. 1C). This indicates a very small structural stabilization of the transcript at volume-occupied solvent conditions. Since Mg^{2+}-cations are known to drive the structural stabilization of RNAs we wondered whether any larger impact of the crowding agent was masked by the presence of Mg^{2+}-cations. As a consequence, we re-analyzed the melting profile of RPS12-RNA in the absence of Mg^{2+}. Representative normalized UV-melting curves and their 1st-derivatives ($\partial A_{260}/\partial T$)

Table 1. A: Helix/coil transition temperatures (T_m) of the *T. brucei* RPS12 pre-mRNA derived from the denaturation profiles shown in Fig. 1. Macromolecular crowding was mimicked by 6% (w/w) PEG_{4000}. B: Summary of SHAPE-derived Gibb's free energies (ΔG) and structural characteristics of the RPS12 transcript in dilute and crowded solvent conditions. In addition, mean and median SHAPE-reactivities are provided. C: Structural features of the RPS12 pre-mRNA in crowded and/or Mg^{2+}-free solvent conditions compared to dilute solvent conditions at 2mM Mg^{2+}. ΔΔG: difference Gibb's free energies. sens: sensitivity. ppv: positive predictive value. For a comprehensive comparison see to Table S1.

A

	dilute	crowded	dilute -Mg^{2+}	crowded -Mg^{2+}
T_m (°C)$_{main\ transition}$	44.0	44.1	31.0	33.1

B

$\Delta G_{37°C}$ (kcal/mol)	-127.4	-131.3	-91.1	-91.8
fraction ss	0.38	0.38	0.43	0.46
fraction ds	0.62	0.62	0.57	0.54
mean reactivity	0.45	0.39	0.57	0.61
median reactivity	0.35	0.32	0.35	0.37
pseudoknotted	yes	yes	no	no

C

	crowded *vs.* dilute	dilute -Mg^{2+} *vs.* dilute	crowded -Mg^{2+} *vs.* dilute
ΔT_m (°C)	<0.5°C	-13.0	-11.0
$\Delta\Delta G_{37°C}$ (kcal/mol)	-3.9	36.3	35.6
sens (%)	100	56	52
ppv (%)	100	60	58

are shown in Fig. 1A. As expected, at dilute solvent conditions the melting profiles changed drastically: The T_m of the main transition shifted by 13°C from 44°C to 31°C and the formerly high temperature transition at 78°C disappeared alltogether. However, as before, the PEG_{4000}-induced stabilization was very small with a ΔT_m of 2°C for the main transition (Fig. 1A and Table 1A). This demonstrates that the crowding-driven stabilization of RPS12 RNA is by far weaker than the stabilization by divalent cations and that the impact on the overall structure of the transcript is minute.

As a follow up of these experiments we analyzed the effects of macromolecular crowding by probing the structure of the RPS12 RNA with nucleotide resolution. For that we used selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) (Merino *et al.*, 2005). SHAPE monitors the local nucleotide flexibility of conformationally unrestrained nucleotides. *In toto* 282 nucleotides were interrogated using the four different solvent regimes depicted in Fig. 2. The nucleotide flexibility is measured in normalized SHAPE-units (SU) and as expected, the RPS12 transcript displays a complex reactivity pattern including highly reactive (>0.8SU) and almost unresponsive (<0.35SU) sequence regions in dilute solvent conditions (Leeder *et al.*, 2016). Sequence stretches with moderate (0.35≤SU≤0.8) to high flexibility are mostly clustered and alternate with unreactive sequence stretches (Fig. 2A). About half of the nucleotide positions are scored inflexible and roughly 10% are highly flexible (>0.8SU). The same probing experiments were performed at crowded conditions and the SHAPE-reactivity differences in the two solvent regimes were plotted as a difference ($\Delta_{crowded\text{-}dilute}$) SHAPE profile (Fig. 2B). The two data sets are characterized by Pearson (*r*) and Spearman (ρ) correlation coefficients ≥0.92 (Table S1C/D) indicating that the structure of the RPS12 RNA is nearly identical in the two solvent regimes.

As expected, a comparison of the SHAPE-profiles in the absence of Mg^{2+}-ions resulted in a different picture. In the absence of the divalent cation both ΔSHAPE-profiles change throughout the entire primary sequence (Fig. 2C/D). About 20% of the nucleotides display reactivity-changes >|0.2|SU and an additional 33% show more than |0.35|SU difference. As a result the data sets only correlate with correlation coefficients of *r*=0.27 (ρ=0.35) (crowded-Mg^{2+} *vs.* dilute) and *r*=0.4 (ρ=0.51) (dilute-Mg^{2+} *vs.* dilute) (Table S1C/D). This behaviour translates to the 2D-structure level when the normalized SHAPE-reactivities are used as pseudo free energy constraints to guide the structure prediction. Fig. 3 shows the pseudoknotted 2D-structure of the RPS12 transcript in dilute and crowded solvent conditions. Molecular crowding has no effect on the transcript if Mg^{2+}-ions are present. However, since the average and median nucleotide flexibility are slightly

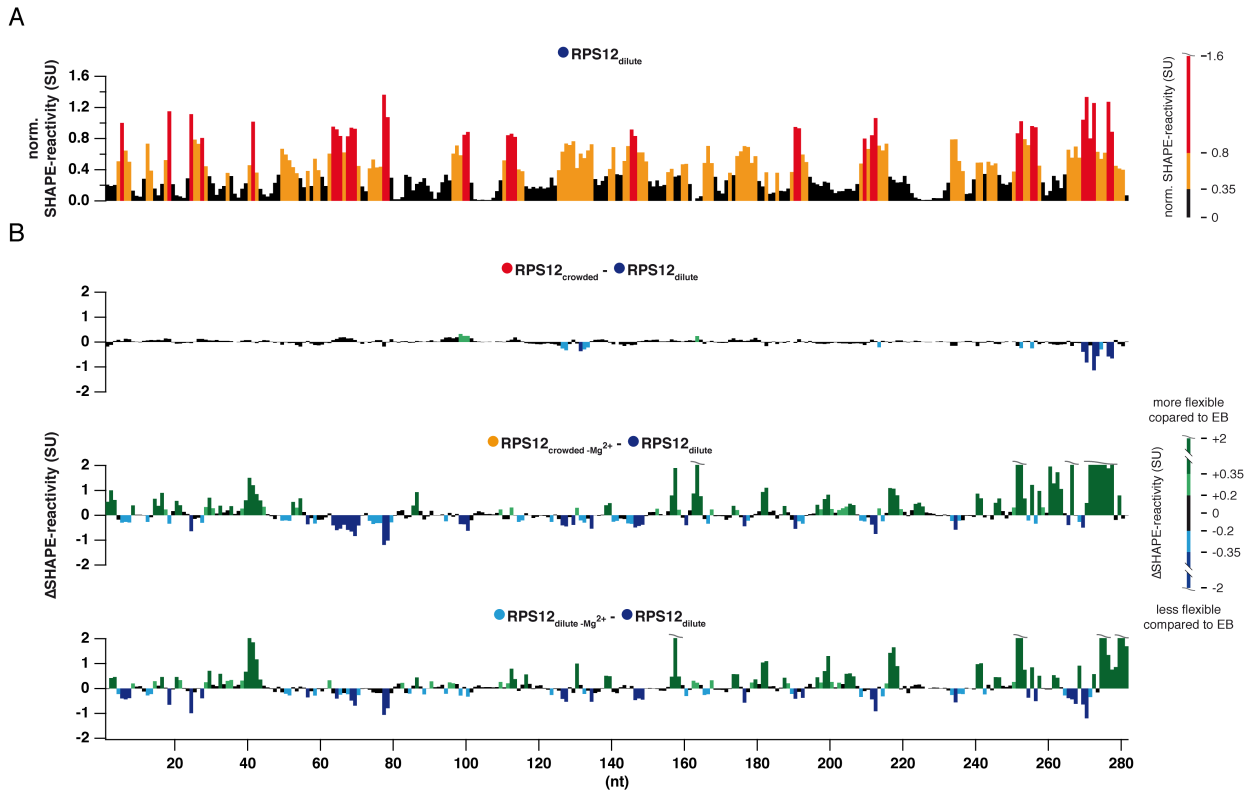


Figure 2. A: Normalized SHAPE-reactivity profiles of the RPS12 pre-mRNA in dilute solvent conditions. Black: low (<0.35SU); orange: medium (0.35≤SU<0.8); red: high (≥0.8) normalized SHAPE-reactivities. SU: SHAPE-unit. nt: nucleotide(s). B: Difference (Δ) SHAPE-reactivity profiles of the RPS12 transcript at crowded and/or Mg^{2+}-free solvent conditions. Green: nt-positions with increased SHAPE-reactivities (dark green: 0.35≤SU≤2; light green: 0.2≤SU<0.35). Blue: nt-positions with decreased SHAPE-reactivities (light blue: -0.2≥SU<-0.35; dark blue: -0.35≥SU≥-0.2. Black: non-responsive nt-positions (-0.2≥SU≤0.2).

decreased at crowded solvent conditions this results in a decrease of the Gibbs free energy (ΔG) of −3.9kcal/mol (Tab. 1B/C). By contrast, in the absence of Mg^{2+}-ions the two structures are characterized by a roughly 15% higher ΔG and the absence of the pseudoknot. Sixty percent of the nucleotides retain their structural context at all conditions studied (Fig. S2).

These results are in line with the published data of Soto *et al.*, 2007 and Tyrrell *et al.*, 2015. As expected, Mg^{2+}-ions exert a stabilizing effect on the global fold of RNA molecules, which is reflected in a ΔT_m of 15°C and a ΔΔG of −36kcal/mol. The contributions of PEG-induced macromolecular crowding to the RNA-stability were less pronounced. Crowding conditions caused an increase of the main T_m≤2°C and a decrease of the Gibb's free energy of maximally 4kcal/mol. Similar trends have been described by Katari *et al.*, 2013 and by Kilburn *et al.*, 2013. The results obtained by SHAPE- and UV-thermal melting experiments in the absence of Mg^{2+}-ions are at a first glance contradictory. The addition of PEG_{4000} resulted in an increase of the T_m by about 2°C but had no effect on the ΔG. Crowding conditions on average even increased the local nucleotide flexibility and had a destabilizing effect on some formerly stable structural elements (Table 1 and Fig. 3). This can be explained by the excluded volume effect lowering the degrees of freedom for conformational sampling ultimately trapping the RNA in misfolded states (Kilburn *et al.*, 2013; Tyrrell *et al.*, 2015). When Mg^{2+} is present, charge compensation by the ion-sphere and/or chelated Mg^{2+}-ions lead to an electrostatic collapse of the RNA thereby generating a highly compact folding state that is further

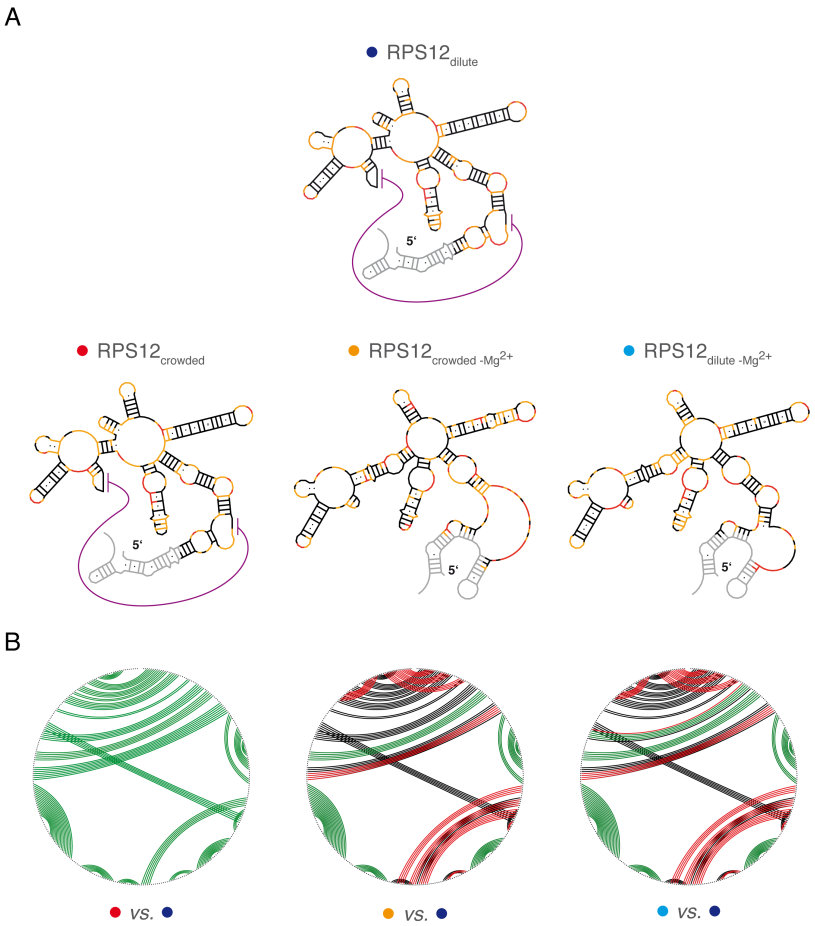


Figure 3. A: SHAPE-derived minimum free energy structures (MFE)-2D-structures of the RPS12 pre-mRNA - top: at dilute solvent conditions in the presence of 2mM Mg^{2+}; bottom: at crowded solvent conditions in the presence/absence of 2mM Mg^{2+}. B: Circle-plot comparison of SHAPE-derived 2D-structures of the RPS12 pre-mRNA. Base pairs (bp) are shown as coloured lines. Green: bp present in both, dilute and crowded solvent conditions. Black: bp unique to dilute solvent conditions in the presence of Mg^{2+}. Red: bp unique to the specific solvent conditions.

stabilized by the excluded volume effect (Heilman-Miller *et al.*, 2001a; Heilman-Miller *et al.*, 2001b; Thirumalai *et al.*, 2001; Grilley *et al.*, 2006; Soto *et al.*, 2007; Kilburn *et al.*, 2010; Leipply & Draper, 2010; Leipply & Draper, 2011). In the presence and absence of Mg^{2+}-ions the crowding-induced stabilization of the RPS12 pre-mRNA over the entire temperature range most likely results from a perturbed folding landscape due to a reduced conformationally freedom (Kilburn *et al.*, 2013). Based on the data we conclude that there is no difference in the 2D-structure of RPS12 pre-mRNA in dilute and crowded solvent conditions at physiological Mg^{2+}-ion concentrations. This is in line with a recent study of Tyrrell *et al.*, 2015. The authors used SHAPE to compare the folding of the aptamer domain of the adenine riboswitch at dilute, crowded and *in vivo* conditions in its free and ligand-bound state. They demonstrated that the 2D-structure of the ligand-free aptamer is recapitulated correctly at all conditions despite less pronounced 3D-interactions *in vitro*. Furthermore, ligand addition resulted in the formation of a more compact conformation involving several higher order 3D-interactions that were invariantly mapped in dilute and crowded solvent conditions and inside the cell. The RPS12 transcript displays non-varying local nucleotide flexibilities as well as identical melting profiles at dilute and crowded conditions given Mg^{2+} is present. This suggests a highly stable and compact structure that obviously is not affected by volume exclusion effects. Thus, we assume that the *in vitro* experiments properly reproduce the *in vivo* situation, at least on the level of RNA secondary structure.

Materials and methods

Cloning and RNA synthesis. The mitochondrial gene encoding ribosomal protein S12 (RPS12) was PCR-amplified from *T. brucei* Lister 427 genomic DNA (Cross, 1975) using the following DNA-oligonucleotide primers (KpnI and SacI restriction endonuclease recognition/cleavage sites are underlined): RPS12_forw. GGGGTACCCTAATACAC-TTTTGATAACAAACTAAAG; RPS12_rev. CCGAGCTC-CCTACCAAACATAAATGAACCTG. The PCR-amplicon

was cloned into the KpnI and SacI endonuclease restriction sites of phagemid pBS SKII$^-$ (Invitrogen) and transcripts were generated by run off *in vitro* transcription from linearized plasmids using T7-RNA polymerase. RNAs were purified from non-incorporated NTP's by size exclusion chromatography, EtOH-precipitated and dissolved in 10mM Tris/HCl pH 7.5, 1mM EDTA (TE).

UV-melting curves. RPS12-transcripts were dissolved in 0.5xTE pH7.5 (50µL), heated to 95°C (2min) and snap cooled on ice before the addition of a concentrated folding buffer to yield a final volume of 0.5mL. Final buffer concentration were: 5mM Na-cacodylate pH6.8, 70mM NaCl and 2mM $MgCl_2$ or 5mM Na-cacodylate pH6.8 and 30mM NaCl (-Mg^{2+}). Volume occupied conditions were generated by supplementing the folding buffer with PEG_{4000} to yield a final concentration of 6% (w/w). This concentration exceeds the polymer crossover concentration $\phi^*_{PEG4000}$=4% (w/w), which marks the transition from a semi-dilute to a crowded solvent regime. RNA concentrations were adjusted to A_{260}=0.5. Denaturation/renaturation profiles were measured at 260nm between 20°C and 95°C at a heating/cooling rate of 1°C/min (data acquisition: 0.3 data points/°C). Melting temperatures (T_m) were obtained from the maximum of the first derivatives ($\partial A_{260}/\partial T$) of the melting curves.

SHAPE-modification. The modification reagent 1–methyl–7–nitroisatoic anhydride (1M7) was synthesized as described (Turner *et al.*, 2013). RPS12 pre-mRNA (0.1µM) was denatured by heating to 95°C (2min) followed by snap cooling on ice. RNA refolding was achieved by equilibration in 20mM HEPES pH7.5, 30mM KCl, 10mM $MgCl_2$ (dilute sovent conditions) or 20mM HEPES pH7.5 and 30mM KCl (-Mg^{2+}) for 30min at 27°C, which represents the optimal growth temperature of insect-stage trypanosomes. Crowding conditions were induced by a PEG_{4000}-concentration of 6% (w/w). RNA samples were split and treated either with 3.5mM 1M7 in DMSO or the same volume of neat DMSO. Modification reactions were quenched after 70sec by the addition of an equal volume of water. RNAs were recovered by EtOH precipitation and desalted by size exclusion chromatography.

Reverse transcription and data processing. Equimolar amounts of fluorescently labelled DNA oligonucleotide primer T3 reverse: 6-FAM/JOE/TAMRA-AATTAACCCTCACTAAAGGGAAC were annealed to 1M7-modified or unmodified RNA samples in 0.25xTE pH7.5 by heating to 95°C (2min), cooling to 50°C (10min) and snap cooling on ice. Reverse transcription was performed in 50mM Tris/HCl pH8.3, 75mM NaCl, 3mM $MgCl_2$, 2.5mM DTT, 0.25mM each dNTP and 0.75U/mL RiboLock RNase inhibitor (Invitrogen). The reaction was started by prewarming the samples for 1.5min prior to the addition of 5U/mL SuperScript III reverse transcriptase (Invitrogen). RPS12 pre-mRNA was reverse transcribed for 20min at 40°C. Sequencing reactions were carried out using unmodified RNA, fluorescently labelled DNA oligonucleotide primer and ddCTP or ddGTP at a final concentration of 0.125mM each. Reverse transcription was stopped by snap cooling and the addition of 0.1 volume of 4M NaOH followed by heating to 95°C (5min). Samples were pooled, EtOH precipitated and redissolved in HiDi® formamid (ABI/Life technologies) for capillary electrophoresis. Raw electrophoretic traces were analysed using SHAPEfinder (Vasa *et al.*, 2008) utilizing the boxplot approach to determine the number of statistical outliers. Normalized SHAPE-reactivities were the result of averaging a minimum of 3 independent experiments.

SHAPE-directed RNA folding. Normalized SHAPE-reactivities were used as pseudo Gibbs free energy (ΔG) values to guide the folding of the RPS12 pre-mRNA. The minimum free energy (MFE) 2D-structure of the RPS12 transcript was folded with ShapeKnots (Hajdin *et al.*, 2013) included in RNAstructure 5.6 (Reuter & Mathews, 2010). The calculations were performed utilizing the default parameters of RNAstructure and ShapeKnots (m=1.8kcal/mol; b=0.6kcal/mol; p1=0.35kcal/mol; p2=0.65kcal/mol). Structure comparison was performed using CircleCompare, which is included in RNAstructure 5.6. MFE-structures were compared in terms of sensitivity (sens=fraction of bp in the reference structure also present in the non-reference structure) and positive predictive value (ppv=fraction of bp in the non-reference structure also occurring in the reference structure). RPS12 pre-mRNA probed in dilute buffer at physiological *i.e.* 10mM Mg^{2+}-ion concentrations served as reference state if not indicated otherwise.

References:

Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nat Chem Biol. 5:593-599.

Chebotareva NA, Kurganov BI and Livanova NB. 2004. Biochemical effects of molecular crowding. Biochem. (Moscow) 69:1239-1251.

Cross GA. 1975. Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. Parasitology. 71:393-417.

De Gennes P. 1979. Scaling concepts in polymer physics. Cornell University Press, Ithaca, NY.

Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. 2014. *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature. 505:696-700.

Doudna JA, Cech TR. 2002. The chemical repertoire of natural ribozymes. Nature. 418:222-228.

Draper DE. 2013. Folding of RNA tertiary structure: Linkages between backbone phosphates, ions, and water. Biopolymers. 99:1105-1113.

Draper DE, Grilley D, Soto AM. 2005. Ions and RNA folding. Annu Rev Biophys Biomol Struct. 34:221-243.

Ellis RJ. 2001. Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci. 26:597-604

Fujimoto T, Nakano S, Miyoshi D, Sugimoto N. 2011. The effects of molecular crowding on the structure and stability of g-quadruplexes with an abasic site. J Nucleic Acids. 2011:857149.

Göringer HU. 2012. 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. Annu Rev Microbiol. 66:65-82.

Grilley D, Soto AM, Draper DE. 2006. Mg^{2+}-RNA interaction free energies and their relationship to the folding of RNA tertiary structures. Proc Natl Acad Sci U S A. 103:14003-14008.

Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. 2013. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proc Natl Acad Sci U S A. 110:5498-5503.

Harve KS, Lareu R, Rajagopalan R, Raghunath M. 2010. Understanding how the crowded interior of cells stabilizes DNA/DNA and DNA/RNA hybrids-in silico predictions and *in vitro* evidence. Nucleic Acids Res. 38:172-181.

Heilman-Miller SL, Pan J, Thirumalai D, Woodson SA. 2001a. Role of counterion condensation in folding of the *Tetrahymena* ribozyme. II. Counterion-dependence of folding kinetics. J Mol Biol. 309:57-68.

Heilman-Miller SL, Thirumalai D, Woodson SA. 2001b. Role of counterion condensation in folding of the *Tetrahymena* ribozyme. I. Equilibrium stabilization by cations. J Mol Biol. 306:1157-1166.

Katari VS, van Esdonk L, Göringer HU. 2013. Molecular crowding inhibits U-insertion/deletion RNA editing *in vitro*: consequences for the *in vivo* reaction. PLoS One. 8:e83796.

Kilburn D, Roh JH, Guo L, Briber RM, Woodson SA. 2010. Molecular crowding stabilizes folded RNA structure by the excluded volume effect. J Am Chem Soc. 132:8690-8696.

Kilburn D, Roh JH, Behrouzi R, Briber RM, Woodson SA. 2013. Crowders perturb the entropy of RNA energy landscapes to favor folding. J Am Chem Soc. 135:10055-10063.

Kozer N, Schreiber G. 2004. Effect of crowd-ing on protein-protein association rates: Fun-damental differences between low and high mass crowding agents. J Mol Biol. 336:763-774.

Kozer N, Kuttner YY, Haran G, Schreiber G. 2007. Protein-Protein association in polymer solutions: from dilute to semidilute to concen-trated. Biophys J. 92:2139-2149.

Lambert D, Draper DE. 2007. Effects of osmolytes on RNA secondary and tertiary structure stabilities and RNA-Mg^{2+} interactions. J Mol Biol. 370:993-1005.

Lambert D, Leipply D, Shiman R, Draper DE. 2009. The influence of monovalent cation size on the stability of RNA tertiary structures. J Mol Biol. 390:791-804.

Lambert D, Leipply D, Draper DE. 2010. The osmolyte TMAO stabilizes native RNA tertiary structures in the absence of Mg^{2+}: evidence for a large barrier to folding from phosphate dehydration. J Mol Biol. 404:138-157.

Lee HT, Kilburn D, Behrouzi R, Briber RM, Woodson SA. 2015. Molecular crowding overcomes the destabilizing effects of mutations in a bacterial ribozyme. Nucleic Acids Res. 43:1170-1176.

Leeder WM, Voigt C, Brecht M, Göringer HU. 2016. The RNA chaperone activity of the *Trypanosoma brucei* editosome raises the dynamic of bound pre-mRNAs. Sci Rep. 6:19309.

Leipply D, Draper DE. 2010. Dependence of RNA tertiary structural stability on Mg^{2+} concentration: interpretation of the Hill equation and coefficient. Biochemistry. 49:1843-1853.

Leipply D, Draper DE. 2011. Evidence for a thermodynamically distinct Mg^{2+} ion associated with formation of an RNA tertiary structure. J Am Chem Soc. 133:13397-13405.

Manning GS. 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. Q Rev Biophys. 11:179-246.

McGinnis JL, Liu Q, Lavender CA, Devaraj A, McClory SP, Fredrick K, Weeks KM. 2015. In-cell SHAPE reveals that free 30S ribosome subunits are in the inactive state. Proc Natl Acad Sci U S A. 112:2425-2430.

Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc. 127:4223-4231.

Minton AP. 2001. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. J Biol Chem. 276:10577-10580.

Misra VK, Draper DE. 2000. Mg(2+) binding to tRNA revisited: the nonlinear Poisson-Boltzmann model. J Mol Biol. 299:813-825.

Miyamoto S, Kashiwagi K, Ito K, Watanabe S, Igarashi K. 1993. Estimation of polyamine distribution and polyamine stimulation of protein synthesis in *Escherichia coli*. Arch Biochem Biophys. 300:63-68.

Miyoshi D, Nakao A, Sugimoto N. 2002. Molecular crowding regulates the structural switch of the DNA G-quadruplex. Biochemistry. 41:15017-15024.

Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics. 11:129.

Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. Nature. 505:701-705.

Soto AM, Misra V, Draper DE. 2007. Tertiary structure of an RNA pseudoknot is stabilized by "diffuse" Mg^{2+} ions. Biochemistry. 46:2973-2983.

Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. 2013. RNA SHAPE analysis in living cells. Nat Chem Biol. 9:18-20.

Srere PA. 1980. The infrastructure of the mitochondrial matrix. Trends Biochem Sci. 5:120–121.

Thirumalai D, Lee N, Woodson SA, Klimov D. 2001. Early events in RNA folding. Annu Rev Phys Chem. 52:751-762.

Thirumalai D, Klimov DK, Lorimer GH. 2003. Caging helps proteins fold. Proc Natl Acad Sci U S A. 100:11195-11197.

Trachman III RJ, Draper DE 2013. Comparison of diamine and Mg^{2+} interactions with RNA tertiary structures: similar *vs*. differential effects on the stabilities of diverse RNA folds. Biochemistry. 52:5911-5919.

Turner R, Shefer K, Ares M. Jr.. 2013. Safer one-pot synthesis of the 'SHAPE' reagent 1-methyl-7-nitroisatoic anhydride (1m7). RNA. 19:1857-1863.

Tyrrell J, McGinnis JL, Weeks KM, Pielak GJ. 2013. The cellular environment stabilizes adenine riboswitch RNA structure. Biochemistry. 52:8777-8785.

Tyrrell J, Weeks KM, Pielak GJ. 2015. Challenge of mimicking the influences of the cellular environment on RNA structure by PEG-induced macromolecular crowding. Biochemistry. 54:6447-6453.

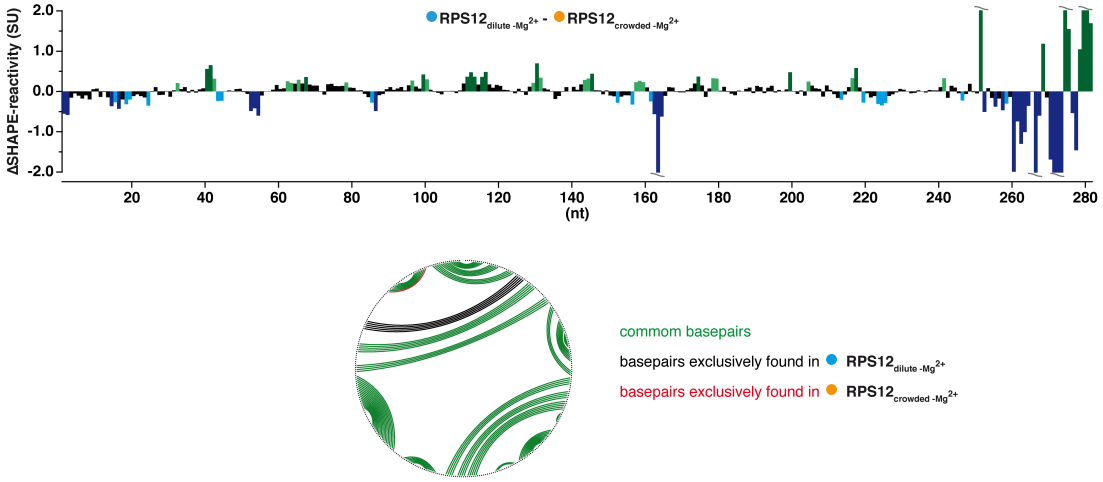
Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. 2008. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA. 14:1979-90.

Zimmerman SB, Trach SO. 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. J Mol Biol. 222:599-620.

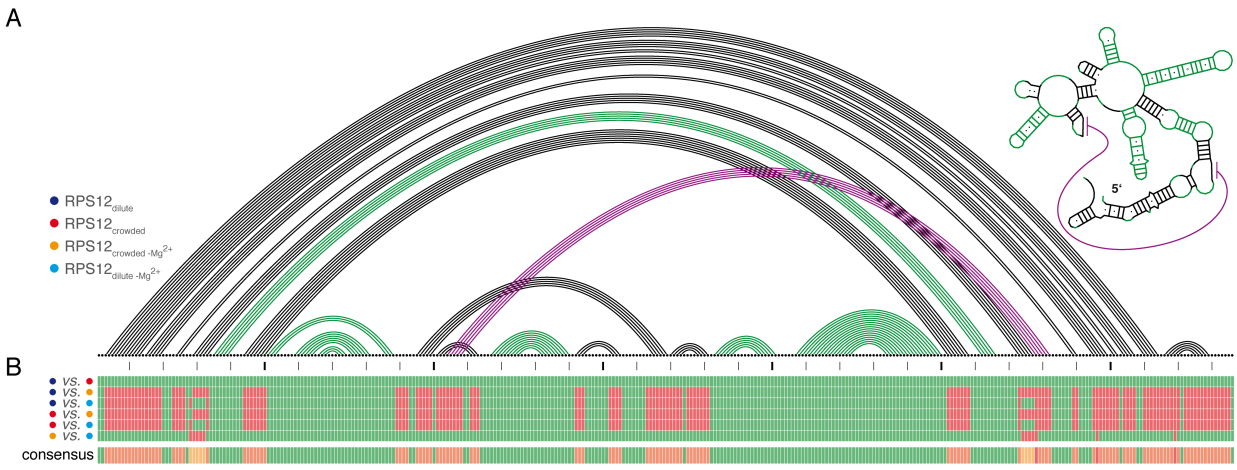
Author Contributions:
Conceived and designed experiments: HUG, WML
Performed experiments: WML, SV

Supplementary Tables/Figures

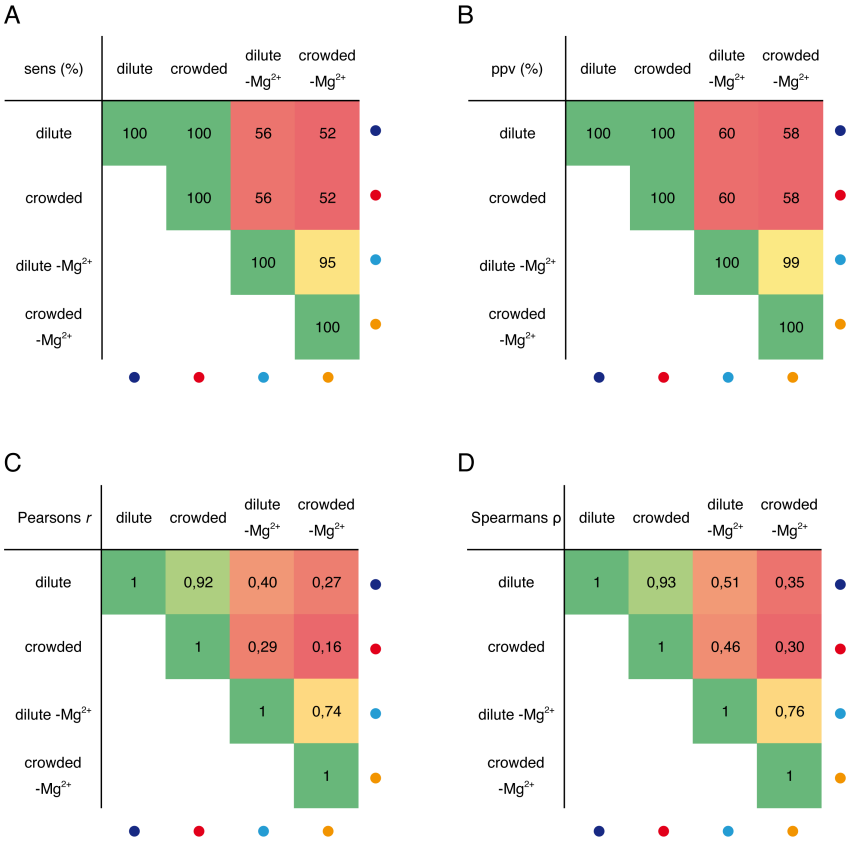
Supplementary Figure S1. Difference SHAPE-reactivities of the RPS12 pre-mRNA in dilute and crowded solvent conditions in the absence of Mg^{2+}. The underlying circleplot visualizes common and mutually exclusive bp in the two MFE-2D-structures.



Supplementary Figure S2. A: Arc-representation of the bp-pattern in the RPS12 pre-mRNA at dilute solvent conditions. Black: variable bp occurring in a subset of conformations. Green: invariable bp common to all 2D-structures. Purple: pseudoknotted bp (exclusive to the presence of Mg^{2+}). The corresponding 2D-structure is shown in the upper right. B: Array comparing the different 2D-structures. Pairwise comparisons are are colour coded as in Fig. 1A. Green: invariant nt; red: nt in a different context (*i.e.* single-stranded *versus* double-stranded or differently paired). Consesus is comparing pairwise 2D-structure comparison. Prevalence of a nucleotide being in an invariant or variable structural context is colour coded. Green: invariant nt; yellow: nt rarely in a variable context; orange: nt prevalently in a variable context; red: nt largly in a variable context compared to dilute buffer conditions in presence of 10mM Mg^{2+}.

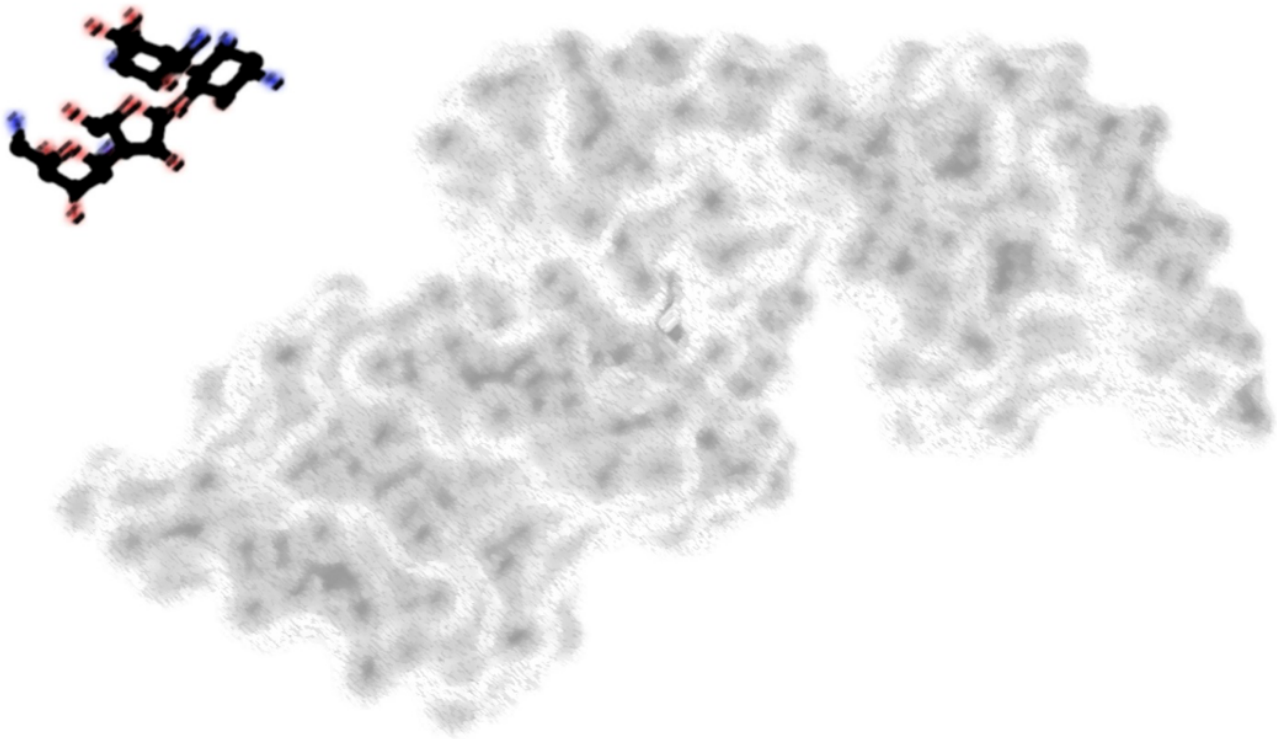


Supplementary Table S1. Matrices comparing the SHAPE-derived modification data of the RPS12 pre-mRNA on the single nt (A/B) and 2D-structure level (C/D). Pearson (*r*) and Spearman (ρ) correlation coefficients were used to compare the different SHAPE-profiles while sensitivity (sens) and positive predictive values (ppv) were applied to compare the SHAPE-derived 2D-structures.



Chapter IV

Charge Reduction and Thermodynamic Stabilization of Substrate RNAs Inhibit RNA Editing



Abstract

African trypanosomes cause a parasitic disease known as sleeping sickness. Mitochondrial transcript maturation in these organisms requires a RNA editing reaction that is characterized by the insertion and deletion of U-nucleotides into otherwise non-functional mRNAs. Editing represents an ideal target for a parasite-specific therapeutic intervention since the reaction cycle is absent in the infected host. In addition, editing relies on a macromolecular protein complex, the editosome, that only exists in the parasite. Therefore, all attempts to search for editing interfering compounds have been focused on molecules that bind to proteins of the editing machinery. However, in analogy to other RNA-driven biochemical pathways it should be possible to stall the reaction by targeting its substrate RNAs. Here we demonstrate inhibition of editing by specific aminoglycosides. The molecules bind into the major groove of the gRNA/pre-mRNA editing substrates thereby causing a stabilization of the RNA molecules through charge compensation and an increase in stacking. The data shed light on mechanistic details of the editing process and identify critical parameters for the development of new trypanocidal compounds.

Introduction

Infections with protozoal pathogens have a global impact, which in part is reflected in the long-standing search for antiprotozoal compounds. Unfortunately, effective treatments for the different diseases are by and large not available (Stuart *et al*., 2008). This holds especially true for African trypanosomiasis a parasite infection also known as African sleeping sickness. The disease is a medical problem in many parts of sub-Saharan Africa due to the lack of effective therapeutics and an increasing resistance of the parasite to long-established chemotherapeutics (Steverding, 2010). Causative agent of sleeping sickness is the organism *Trypanosoma brucei* - an extracellular parasite that multiplies in the blood and the lymphatic and cerebrospinal fluids. Some progress in the development of new trypanocidal compounds has recently been made by reformulating existing drugs while at least two other compounds are currently being tested in clinical trials (Stich *et al*., 2013). In addition, alternative approaches to address the need for safe and effective drugs have been explored for instance by using conjugated parasite-specific nanobodies (Baral *et al*., 2006) or SELEX-derived, trypanosome-specific RNA aptamers (Homann & Göringer, 2001; Lorger *et al*., 2003; Göringer, 2012a).

Despite these efforts, almost no attention has been paid to RNA-mediated biochemical pathways as potential drug targets. Parasite-specific RNAs and/or ribonucleoprotein (RNP) complexes likely represent vulnerable targets that can be exploited with RNA-interacting, small molecule compounds (Verspieren *et al*., 1987; Hermann, 2000; Tor, 2003). Within that context the RNA editing reaction within the mitochondria of trypanososmes is of special importance. The processing reaction is a key pathway in the life cycle of the parasite (Aphasizhev & Aphasizheva, 2014) and the process is catalyzed by a unique high molecular mass protein complex, the 20S editosome (Göringer, 2012b). Moreover, editing involves a class of small, non-coding RNAs, known as guide (g)RNAs, which only exist in the parasite. An editing reaction cycle is characterized by the site-specific insertion and to a lesser degree deletion of exclusively U-nucleotides into otherwise non-translatable mRNAs. The reaction is catalyzed within a single reaction center on the surface of the 20S "protein-only" editosome (Böhm *et al*., 2012). RNA substrates are pre-edited mRNAs and guide RNAs, which provide the specificity for the U-insertion/deletion reaction.

Although editing is most prominent during the insect life cycle stage of the parasite, it also occurs in the infective bloodstream stage and thus can be considered an ideal drug target (Schnaufer *et al*., 2001). However, only a limited number of inhibiting compounds have been identified to date (Salavati *et al*., 2011). This includes inorganic pyrophosphate (pp_i), which is generated from UTP during the U-insertion reaction and as a consequence inhibits the terminal uirdylyltransferase (TUTase) of the editosome by feedback inhibition. Chemically inert cosolutes such as polyethylenglycol (PEG) have recently been shown to block the reaction indicating that *in vitro* RNA editing is sensitive to molecular crowding conditions (Katari *et al*., 2013). Other inhibitors have been identified by a combined virtual screening/molecular dynamic simulation approach (Demir *et al*., 2014) or by high throughput screening of chemical libraries. This identified compounds such as mitoxanthrone, protoporphyrin IX and D-sphingosine, which inhibit with half-maximal inhibitory concentrations (IC_{50}) in the low micromolar range and likely function by binding to the editosome or individual proteins of the catalytic complex (Liang & Connell, 2010). Similarly, several

naphthalene-derivatives have been identified as low micromolar inhibitors of the editing core enzyme REL1 (RNA editing ligase 1). They interfere with the deadenylation of the enzyme in addition to blocking the integrity and/or assembly of the editosome (Amaro *et al*., 2008; Durrant *et al*., 2010; Moshiri *et al*., 2011).

Although, the 0.8MDa editosome presents a large drug-binding landscape, conceptually it should also be possible to modulate the processing reaction by RNA substrate-binding. Typical examples for RNA-binding, small molecule inhibitors are some antibiotics such as the aminoglycosides, which interfere with protein biosynthesis by directly binding to the ribosomal decoding site of the small subunit ribosomal RNA (Carter *et al*., 2000). Although aminoglycosides show selectivity in their binding of RNA over DNA, they are rather nonselective towards different RNA molecules. Aminoglycosides have been identified to interact and inhibit a wide range of unrelated RNAs and biochemical processes including protein synthesis, group I intron splicing, RNA catalysis as well as viral RNAs (Tor, 2003). This promiscuity is the result of a general electrostatic binding modality and a conformational adaptability of the aminosugars (Hermann, 2002), which bind into the major groove of A-form RNA duplex elements as well as major grooves that are widened by the proximity of a loop or a bulge (Jin *et al*., 2000; Varani *et al*., 2000; Xi *et al*., 2011).

Here we show that both, U-insertion- and U-deletion editing can be inhibited by specific aminoglycosides with IC_{50}-values in the low micromolar range. The inhibition is a consequence of high affinity binding of multiple aminoglycoside molecules into the major groove of the gRNA/pre-mRNA substrate RNAs, which results in a thermodynamic stabilization of the RNAs. The data shed light on mechanistic details of the editing reaction cycle and identify critical parameters for the development of new trypanocidal compounds.

Results

Aminoglycosides inhibit *in vitro* RNA editing. Based on the published data on the inhibition of translation, pre-mRNA splicing and other RNA-driven reactions (Tor, 2003) we tested the ability of a panel of antibiotics including several aminoglycosides to interfere with U-insertion/deletion-type RNA editing. Together we analyzed four 4,5-linked 2-deoxystreptamines (2-DOS), five 4,6-linked 2-deoxystreptamines, three aminocyclitols and three macrolides (for chemical structures see S1 Figure Supporting Information). The two different editing reactions (U-insertion/U-deletion) were tested independently using "precleaved" *in vitro* RNA editing assays (Igo *et al*, 2000, Igo *et al*, 2002). Both assays rely on short, synthetic pre-mRNA/gRNA hybrid RNAs, which are converted into edited products either by the insertion of 3 U-nucleotides (nt) or the deletion of 4 U's (Figure 1A/B). The reaction is catalyzed by 20S editosomes and was performed at varying antibiotic concentrations. RNA reactants, intermediates, edited products as well as non-productive side products of the reaction were electrophoretically separated and densitometrically quantified. Figure 1C/D show two representative experiments using the aminoglycoside neamin. Neamin inhibits both editing reactions in a dose-dependent manner (S3 Figure Supporting Information). The U-insertion reaction is half-maximally (IC_{50}) inhibited at a concentration of 30±8μM and the IC_{50}-value for inhibiting U-deletion editing is 31±2μM. All data are summarized in Figure 1E. Neomycin B (neo) is the most effective inhibitor with IC_{50}-values of 10±1μM and 13±2μM (U-deletion/U-insertion) followed by neamin (31/30μM), tobramycin (47/34μM) and sisomycin (36/60μM). The macrolides

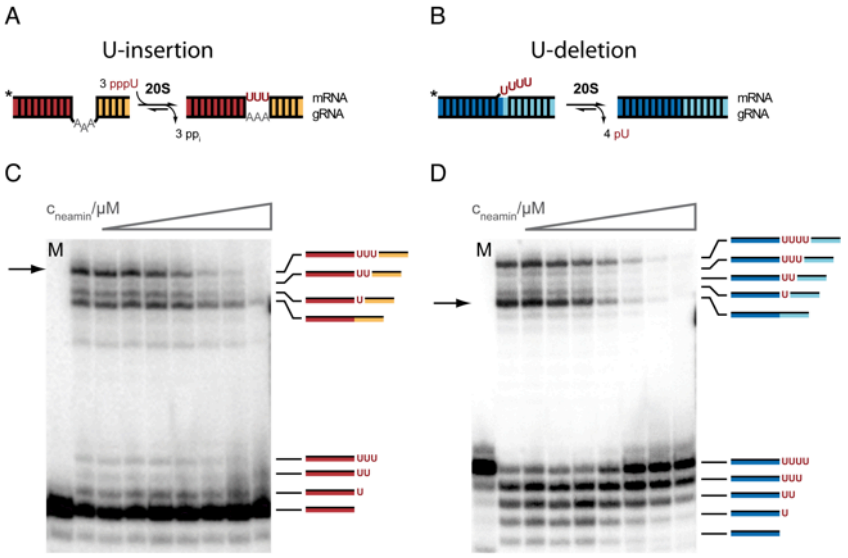


Figure 1. Trypanosome-specific U-insertion and U-deletion RNA editing. Depicted are two gRNA/pre-mRNA pairs emphasizing the helical domains of the U-insertion RNA substrate in red and yellow (A) and the helices of the U-deletion RNA in dark and light blue (B). *In vitro* U-insertion monitors the insertion of 3 U's; *in vitro* U-deletion the removal of 4 U nt. The reaction is catalyzed by the 20S editosome. Primary sequences of the individual RNAs are given in S2 Figure Supporting Information. Inhibition of U-insertion (C) and U-deletion RNA editing (D) by neamin. Radioactively labelled (5'-^{32}P) gRNA/pre-mRNA substrate RNAs were incubated with 20S editosomes in the presence of increasing concentrations of neamin (1.6µM-1.7mM, left to right). RNA reactants, intermediates and edited products (annotated to the right of the two gels) were electrophoretically separated and densitometrically quantified to yield dose response curves (S3 Figure Supporting Information) from which IC_{50}-values were derived (E). Errors are standard deviations (s.d.). M: mock treated sample. *: position of the radioactive label (^{32}P). Arrows indicate the position of the fully edited mRNA products.

E

	compound	IC_{50} (µM) insertion RNA-editing	IC_{50} (µM) deletion RNA-editing
	neamin	30 ± 8	31 ± 2
4,5-2-DOS linked aminoglycosides	neomycin B	13 ± 2	10 ± 1
	paromomycin	173	137
	ribostamycin	325	483
	lividomycin	97	263
4,6-2-DOS linked aminoglycosides	tobramycin	34 ± 4	47 ± 6
	kanamycin A	275	581
	kanamycin B	180	137
	sisomicin	60 ± 5	36 ± 2
	geneticin	430	649
other aminocyclitols	apramycin	287	309
	streptomycin	-	195
	hygromycin B	-	>1000
macrolide antibiotics	erythromycin	-	>100
	azithromycin	-	>100
	spiramycin	-	>100

erythromycin, azithromycin and spiramycin and the aminocyclitols apramycin, streptomycin and hygromycin B showed no effect (IC_{50}≥100µM).

Importantly, the inhibitory effect shows pleiotropic characteristics: next to inhibiting the formation of the fully edited pre-mRNA other steps of the processing cycle are affected as well. In the case of the U-deletion, the exoUase trimming reaction at the 3'-end of the 5'-pre-mRNA fragment is inhibited, although to a lesser degree. Furthermore, religation of the "faithfully" edited reaction product is more sensitive to the aminosugars in comparison to the formation of the unedited ligation side-product. However, while the formation of edited products in both cases is completely stalled at concentrations ≥100µM, the inhibition of the exoUase is not. Lastly, we also analyzed the endonuclease cleavage step of the editing cycle. mRNA cleavage is similarly inhibited in a dose dependent manner and is completely abolished at ≥30µM neomycin B (S4 Figure Supporting Information).

Aminoglycosides bind to editing substrate RNAs. Aminoglycosides have been shown to execute their inhibitory effects through a promiscuous RNA binding activity (Hermann, 2002). Therefore, we analyzed the possible binding of neomycin B (the most potent inhibitor of the editing reaction) to the two editing substrate RNAs using isothermal calorimetry (ITC). As shown in Figure 2A/B neomycin B binds to both, the U-insertion and U-deletion gRNA/pre-mRNA hybrid RNAs with micromolar affinity. The binding is characterized by a macroscopic equilibrium dissociation constants (K_d) of 1.1µM for the U-insertion gRNA/pre-mRNA

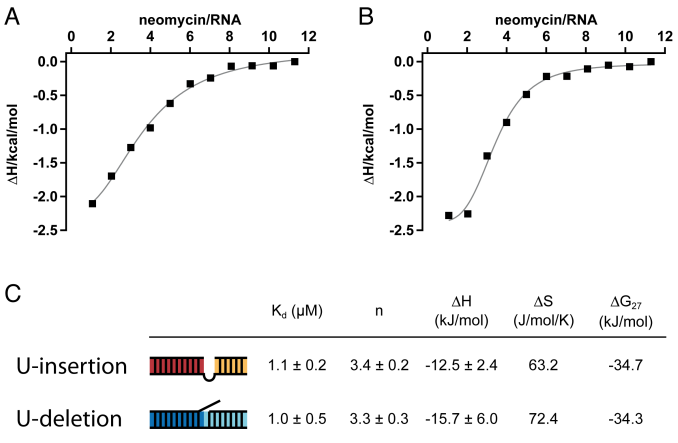


Figure 2. ITC-titration profiles of U-insertion (A) and U-deletion (B) gRNA/pre-mRNA hybrid RNAs with neomycin B. (C) Summary of the derived thermodynamic characteristics of the binding reaction: equilibrium dissociation constant (K_d), number of binding sites (n), enthalpy (ΔH), entropy (ΔS) and Gibbs free energy (ΔG). Errors are standard deviations (s.d.).

hybrid and of 1.0μM for the U-deletion RNA (Figure 2C). The binding reaction is exothermic and at saturation 3 RNA binding sites are occupied with neomycin B. The interaction is enthalpically driven (Jin *et al.*, 2000) with Gibbs free energies in the range of −35kJ/mol. Assuming that the 3 binding sites are identical, the determined macroscopic K_d's were redefined relative to the progress of saturation using a sequential saturation binding site model (Bou-Abdallah *et al.*, 2005; Le *et al.*, 2013). This allowed the calculation of intrinsic binding constants (K_i) from which the fraction of RNA molecules having 1, 2 or 3 binding sites occupied was determined (S5 Figure Supporting Information). At a neomycin B concentration of app. 12μM half maximal occupancy of all 3 binding sites is achieved in aggreement with the above determined IC_{50}-value.

Importantly, neomycin B has no affinity for 20S editosomes as demonstrated by surface plasmon resonance (SPR) binding experiments. At neomycin B concentrations up to 10-fold above the determined K_d of the RNA/neomycin B complex no binding was observed. Only at a ≥100-fold excess a weak interaction was identified (S6 Figure Supporting Information). Similarly, neomycin B does not affect the binding of substrate RNA to 20S editosomes. Even at a neo concentration 250xK_d no dissociation of bound RNA from the complex can be detected (S7 Figure Supporting Information).

Aminoglycosides stabilize editing substrate RNAs. Based on the observation that the aminoglycoside-induced inhibition of editing might be a result of the high affinity interaction of the drug to the substrate RNAs we decided to analyze the consequence(s) of the RNA/aminosugar interaction. For that we performed UV-melting experiments of gRNA/pre-mRNAs hybrid RNAs in the presence of defined concentrations of neomycin B. Representative melting curves for the U-insertion and U-deletion RNAs are shown in Figure 3A/B. In the absence of neomycin B, both RNAs show two separate helix/coil transitions indicating independent melting events of the two helical domains in both RNAs. T_m-values were derived from 1st-derivative plots and identified melting midpoints at temperatures ranging from 47°C to 65°C. In the presence of 1.3μM neomycin B (1.3-fold above K_d) the individual melting transitions shift to higher temperatures with ΔT_m's between 11-13°C. This indicates a strong stabilization of the helical elements in both RNAs equivalent to ΔΔG-values of −7 to −23kJ/mol. All binding reactions show saturation characteristics, which allowed the determination of the number of neomycin B binding sites per RNA (Figure 3C/D). For the U-insertion RNA we identified 3.3 binding sites and for the U-deletion substrate 3.1 binding sites in agreement with the above-described ITC-measurements. All data are summarized in Figure 3E.

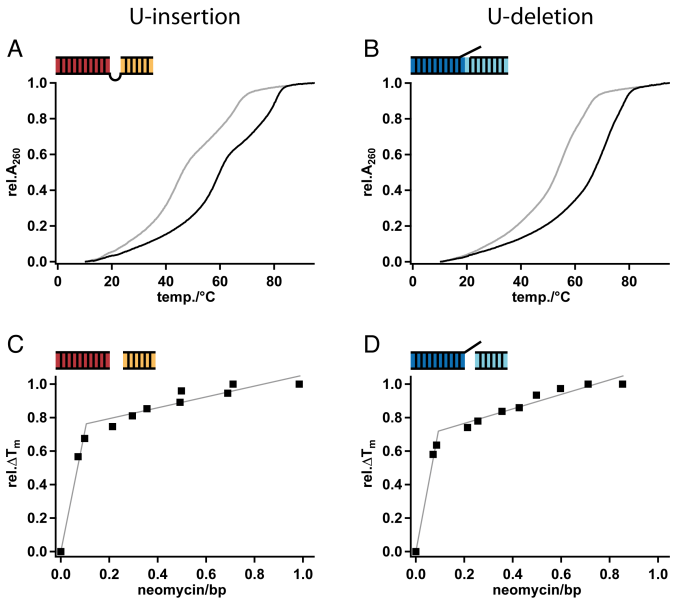
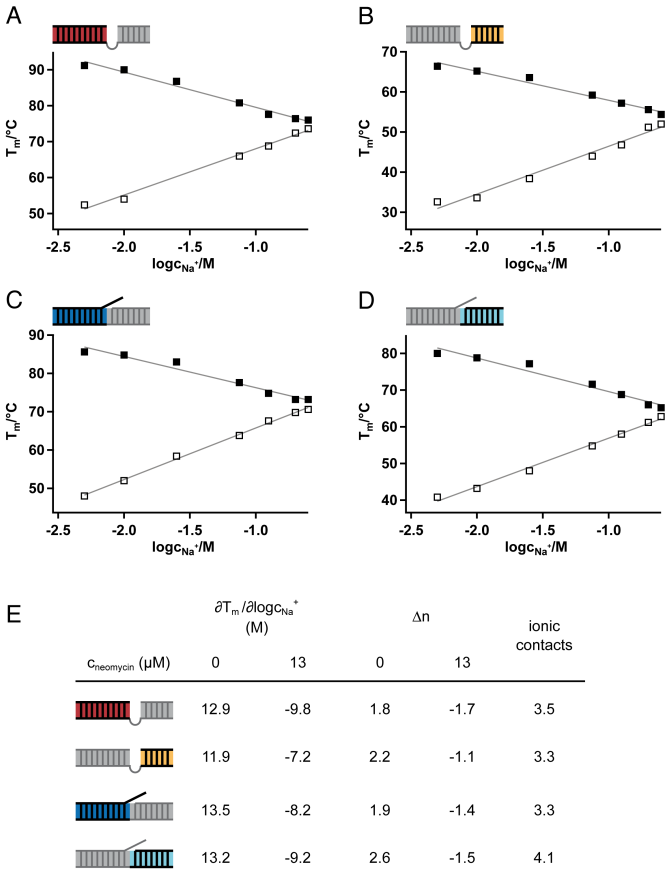


Figure 3. UV-melting profiles of U-insertion (A) and U-deletion (B) gRNA/pre-mRNA substrate RNAs (1µM) in the absence (grey trace) and presence (black trace) of neomycin B. Both RNAs exhibit two distinct helix/coil transitions indicating the independent melting of the two helical domains in both RNAs. In the presence of neomycin B (13µM; 13-fold over K_d) both melting curves shift to higher temperatures indicating a stabilization of the editing RNAs. (C)/(D) Plotting the T_m-value changes of the individual helices as a function of the molar neomycin B/bp ratio results in 3.1 (U-insertion) and 3.3 (U-deletion) neomycin binding sites per editing RNA. (E) Summary of the measured/calculated T_m, ΔT_m and thermodynamic values for the individual helices in both gRNA/pre-mRNA substrate RNAs.

E

	$c_{neomycin}$ (µM)	T_m (°C)	ΔT_m (°C)	ΔH (kJ/mol)	ΔS (J/mol/K)	ΔG_{25} (kJ/mol)	$\Delta\Delta G_{25}$ (kJ/mol)
	0	65.2		-271.3	-672.9	-70.6 ± 3.2	
	1.3	70.1	11.1	-256.4	-621.3	-71.1 ± 2.5	-22.5
	13	76.3		-374.4	-943.4	-93.1 ± 2.4	
	0	47.2		-327.7	-898.8	-60.3 ± 1.5	
	1.3	53.7	11.6	-301.8	-796.9	-64.2 ± 2.3	-6.5
	13	58.8		-283.3	-726.0	-66.8 ± 4.0	
	0	63.8		-279.6	-701.4	-70.5 ± 1.5	
	1.3	69.5	10.6	-262.6	-637.5	-72.5 ± 2.6	-19.3
	13	74.4		-354.9	-889.2	-89.8 ± 2.4	
	0	50.8		-357.5	-976.9	-66.2 ± 3.0	
	1.3	56.1	12.7	-300.3	-782.3	-67.0 ± 2.3	-7.9
	13	63.6		-312.4	-799.0	-74.2 ± 3.3	

The neomycin B/editing substrate RNA interaction relies on ionic contacts. The ability of aminoglycosides to bind to RNA molecules has been attributed to two general phenomena: (i) the conformational adaptability of the sugar molecules and (ii) their "overall" electrostatically-driven RNA binding mode (Hermann, 2002; Tor, 2003). In order to test the contribution of a charge/charge interaction between the polyanionic editing RNAs and the cationic aminoglycosides we analyzed the thermal stability of the individual RNA helices at increasing Na^+-ion concentrations (5-250mM) in the absence and presence of neomycin B. According to polyelectrolyte theory the melting temperature (T_m) varies logarithmically with the salt concentration and the slope of the linear relationship is proportional to the difference in bound counter ions (Δn) in the folded and unfolded RNA. Figure 4A-D summarize the results. In the absence of neomycin B all four T_m=f(log cNa^+) plots are characterized by positive slopes ($\partial T_m/\partial \log cNa^+$) indicating a release of counter ions upon RNA melting. Conversely, in the presence of 13µM neomycin B (13-fold over K_d) the slopes are negative demonstrating an uptake of Na^+-ions during melting. This implies that the temperature-induced strand separation of the individual RNA helices causes the bound aminoglycoside molecules to dissociate thereby freeing negative charges, which are neutralized by an uptake of Na^+-ions. The calculated Δn-values vary from 1.8 to 2.6 in the absence of neomycin B and from −1.1 to −1.7 for the neomycin B-bound RNAs (Figure 4E). Depending on the individual helices this indicates a release of 1 to 2 Na^+-ions upon RNA melting and equally an uptake of 1 to 2 Na^+-ions during melting of the neomycin B-complexed RNAs. The net difference between naked and neomycin B-complexed RNA helices calculates to ~3-4, which implies that in every helical element of the two editing RNAs 3-4 bound Na^+-ions are replaced by 3-4 NH_3^+-groups of the aminoglycoside in order to facilitate the RNA/neomycin B interaction.



E

	$\partial T_m/\partial logc_{Na^+}$ (M)		Δn		ionic contacts
$c_{neomycin}$ (μM)	0	13	0	13	
	12.9	-9.8	1.8	-1.7	3.5
	11.9	-7.2	2.2	-1.1	3.3
	13.5	-8.2	1.9	-1.4	3.3
	13.2	-9.2	2.6	-1.5	4.1

Figure 4. Salt dependence of UV-melting profiles. Sodium ion dependence (10-250mM) of melting temperatures of the U-insertion (A/B) and U-deletion (C/D) gRNA/pre-mRNA editing substrate RNAs (1μM) in the absence (open squares) and presence (filled squares) of neomycin B (13μM). Plots of the melting temperature (T_m) versus log of the Na^+-ion concentration (log cNa^+) for the 4 individual helices of the two editing RNAs. Solid lines: linear regressions of the data points. (E) Summary of the derived data.

Neomycin B binding does not alter the overall structure of the pre-mRNA/gRNA hybrid RNAs. To analyze whether the aminosugar-induced thermodynamic stabilization of editing substrate RNAs is a result of structural changes in the gRNA/pre-mRNA hybrid molecules we performed a conformational analysis using circular dichroism (CD) spectroscopy. Representative spectra are shown in Figure 5A/B. As expected, both RNAs show the characteristics of A-form RNA (Clark *et al.*, 1997): a dominant negative signal at 212nm, a strong positive band at 265nm and a weak negative signal at roughly 300nm (with a crossover point at 224-234nm). The addition of neomycin B at molar ratios up to 40/1 alters both spectra in a concentration-dependent fashion although only moderately. Changes in the ellipticity indicate an increase in stacking (Scheunemann *et al.*, 2010) in line with the formation of a more compact structure of the two RNAs around the ligand binding sites (Zhao *et al.*, 2005). However, the overall A-form geometry of the two RNAs is not altered upon binding of the aminosugar. Plotting the change in ellipticity (ΔΕ) at 273nm as a function of the molar neomycin B/RNA ratio yields in both cases saturation-type titration curves from which, in agreement with the above-described ITC- and UV-melting data, 3.1 (U-insertion) and 3.3 (U-deletion) binding sites were calculated (Figure 5C/D).

Modeling/docking of the neomycin B/editing RNA complexes. In order to gain a structural understanding of the aminoglycosid-mediated inhibition of editing we decided to model the two gRNA/pre-mRNA hybrid RNAs based on the above-described experimental data and to predict possible neomycin B binding sites by ligand docking. The two RNAs were interactively modeled with the help of ERNA3D© 2.0 using standard A-form geometries for the helical elements and maximal stacking of the ss-nucleotides at the editing junction (Niemann *et al.*, 2008). Neomycin B was docked into the two RNAs by blind docking using Autodock Vina 1.1.2 (Ranjan *et al.*, 2010; Trott & Olson, 2010; Hamilton & Arya, 2012). Neomycin B was predicted to bind into the major groove of the individual helices in both RNAs with in each case 9 possible conformations of very similar binding energies (variations between −38 and −41kJ/mol) and orientations. The pairwise root-mean-square deviation (RMSD) between the different conformations was maximally 3.3Å and minimally 0.5Å. Each neomycin B molecule covers a length of

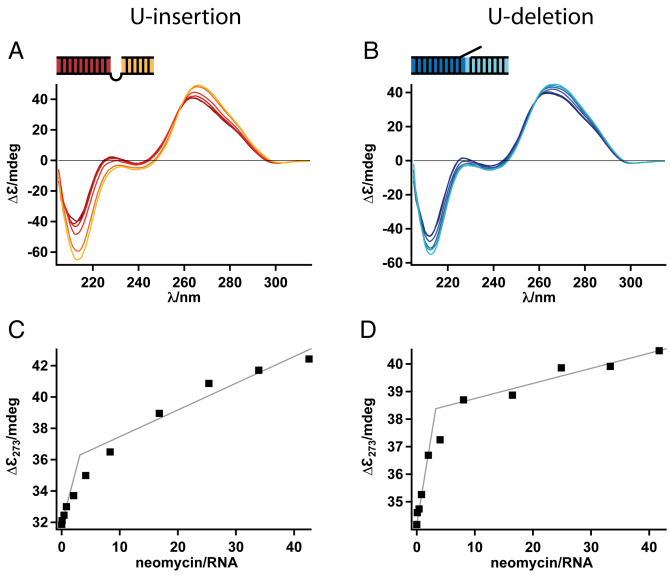


Figure 5. CD-spectra of RNA editing substrate RNAs at a concentration of 12µM. (A) U-insertion gRNA/pre-mRNA hybrid RNA (red trace). (B) U-deletion gRNA/pre-mRNA hybrid RNA (dark blue trace). Both spectra show typical A-form characteristics. Adding increasing concentrations of neomycin B (1.3µM-0.5mM) yields the spectra shown in orange to yellow (U-insertion) and light blue to cyan (U-deletion). (C/D) Plotting the spectral changes at 273nm as a function of the molar neomycinB/RNA ratio results in 3.1 (U-insertion) and 3.3 (U-deletion) neomycin binding sites per editing RNA.

about 6bp, which allows the integration of 2 neo molecules into the larger helical elements in both RNAs and of 1 neomycin B molecule into the shorter stems in agreement with the calculated neomycin B/RNA stoichiometry of 3:1 (Figure 6A/B). All predicted neomycin B conformations show 3-4 ionic contacts to the RNA backbone with an average distance of 3.7Å of an amino-nitrogen of the ligand to an oxygen atom of a phosphate group (Figure 6C). Neomycin B binding does not distort the helical geometries of the two RNAs and a superposition of the six different neo conformations within their binding pockets is shown in Figure 6D.

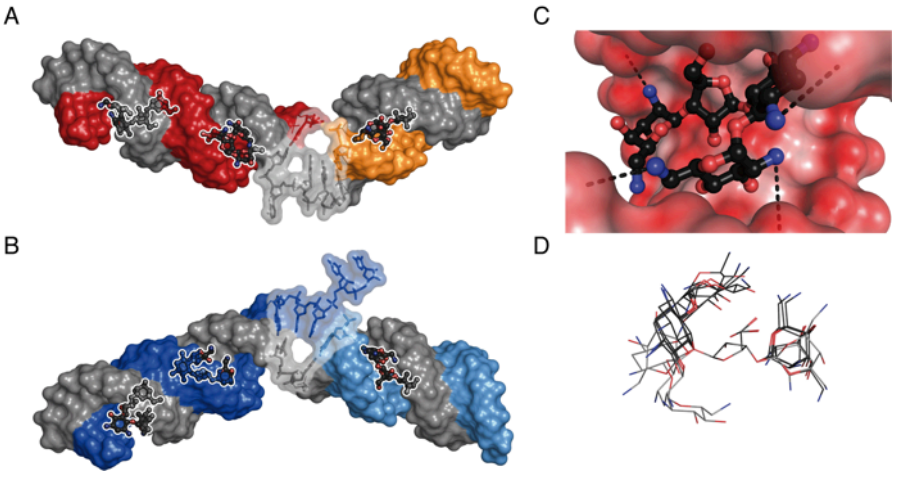


Figure 6. Docking-derived molecular models of neomycin B-complexed U-insertion (A) and U-deletion (B) gRNA/pre-mRNA hybrid RNAs. gRNA molecules are shown in grey; pre-mRNAs in red/yellow (U-insertion) and dark blue/light blue (U-deletion). Neomycin B is shown in a ball and stick representation (nitrogen-blue, oxygen-red, carbon-grey). Three neomycin B molecules are bound into the major groove of the individual RNA helices in line with the results from the ITC-, CD- and UV-melting experiments. The U-insertion gRNA/pre-mRNA hybrid RNA is bent by an angle of 134°, the U-deletion RNA by 138°. Nucleotides at the editing sites are depicted in a transparent envelop. (C) Calculated electrostatic surface potentials (red: negative; white: neutral; blue: positive) of a neomycin B binding site within the major groove of the U-deletion editing substrate. Possible ionic contacts are shown as dashed lines. (D) Conformational superposition of all 6 neomycin B molecules aligned to the ribose moiety of the molecule.

Discussion

The RNA editing reaction within the mitochondria of African trypanosomes represents an ideal drug target for two reasons: First, editing is a required biochemical pathway for the survival of the parasite and second, the reaction cycle involves a multi-enzyme machinery (the editosome) and a class of small, non-coding RNAs (gRNAs) that only exist in the parasite. Although some editosome-interacting compounds have been identified (Salavati *et al.*, 2011), no inhibitor targeting the substrate gRNA/pre-mRNA hybrid RNAs of the processing reaction is known to date.

Here we demonstrate that the aminosugars neamin, neomycin B, tobramycin and sisomycin inhibit RNA editing at micromolar concentrations. These values are in a similar range to the inhibitory concentrations for other RNA-driven biochemical processes such as the neomycin B-induced inhibition of the hammerhead ribozyme (Stage *et al.*, 1995), the inhibition of the human hepatitis delta virus self-cleavage RNA (Rogers *et al.*, 1996)

or RNase P RNA (Mikkelsen *et al*., 1999). The two editing sub-reactions (U-insertion/U-deletion) are repressed with similar qualitative and quantitative characteristics, which is in line with the fact that the two reactions are catalyzed out of a single RNA substrate binding site (Böhm *et al*., 2012). In agreement with observations in other RNA-driven systems we demonstrate that the inhibitory effect is a direct consequence of RNA binding. The aminoglycoside-gRNA/pre-mRNA complexes have micromolar affinities and are characterized by a stoichiometry of 3 aminosugars per gRNA/pre-mRNA hybrid. The RNA binding sites localize to the major groove of the two helical domains of the hybrid RNAs with a ratio of about 1 aminoglycoside/10bp. This agrees with published data in other systems (Chen *et al*., 1997; Varani *et al*., 2000; Xi *et al.*, 2011). The binding reaction represents a quasi "shape readout" of the helical structure of the two gRNA/pre-mRNA hybrid RNAs, which is consistent with NMR and X-ray studies that have shown that the conformational specificity of aminosugars is a result of a set of favorable interactions deep in the major groove of A-form duplex structures (Fourmy *et al*., 1998a; Fourmy *et al*., 1998b; François *et al*., 2005; Zhao *et al*., 2005). Importantly, the binding reaction does not induce a gross structural perturbation of the A-form geometry of the two RNA helices, instead it results in a thermodynamic stabilization in the range of −7 to −23kJ/mol. This could be the result of a charge neutralization between the cationic aminoglycosides and the polyanionic RNAs (Kaul *et al*., 2003; Freire *et al*., 2007), especially since at least three NH_3^+ groups contribute to a single neomycin B-gRNA/pre-mRNA interaction. Within the context of the editing reaction this suggests that the editosome/substrate RNA interaction relies on a defined number of negative charges, which when neutralized cause the processing reaction to stall. This is further supported by the fact that *in vitro* RNA editing is performed at low monovalent cation concentrations (≤30mM) because the reaction is known to be sensitve to high salt conditions.

Furthermore, the data suggest that the processing reaction requires a defined structural flexibility of the gRNA/pre-mRNA hybrid RNAs. Aminosugar binding induces a stabilization or "stiffening" of the substrate RNAs, which leads to inhibition. This scenario is supported by the observation that editosomes are known to tolerate dynamic RNA rearrangements at editing sites simply because the primary sequence of the pre- and partially edited mRNAs is constantly changed as a result of the numerous U-insertions and deletions. An inability to re-organize the structure of the gRNA/pre-mRNA hybrid molecules might therefore stall the reaction. This is supported by the fact that editosomes have been shown to execute RNA chaperone activity to resolve folded RNA motifs (Böhm *et al*., 2012) in addition to the involvement of RNA helicases (Missel *et al*., 1997; Li *et al*., 2011; Kruse *et al*., 2013). Related phenomena have been described in the case of tobramycin, which when bound to tRNAAsp, locks the RNA in a non-functional conformation (Walter *et al*., 2002) and the kissing loop-mediated dimerization of HIV-1 RNA, which is stabilized by the binding of aminoglycosides thereby blocking the kissing loop to duplex conversion (Bernacchi *et al*., 2007)

Importantly, the inhibitory effect is not a result of a direct structural interference of the aminosugar molecules at the editing site. This is supported by our CD and modeling/docking data. Instead, it's a consequence of multiple binding events distal to the RNA editing site, which is transduced to the catalytic domain. A likely canditate for this mediator function might be the water structure of the substrate gRNA/pre-mRNA hybrid RNAs. Electrostatically confined water

molecules in the 1st-hydration shell of RNAs have been shown to add to the plasticity of RNA/ligand interactions (Westhof, 1993; Hermann & Patel, 1999;). Therefore, we suggest that aminoglycoside-induced distal changes of the RNA hydration shell transduce to the editing site to stall the processing reaction indirectly. This hypothesis is supported by the fact that *in vitro* RNA editing is inhibited at crowded solution conditions (Katari *et al*., 2013). Synthetic crowding agents such as polyethylene glycol (PEG) are known to affect the hydration state and as a consequence the functionality of nucleic acid molecules (Koumoto *et al*., 2008; Pramanik *et al*., 2011).

The described results can be used to define a set of criteria to search for improved RNA-interacting editing inhibitors for instance by structure-guided approaches as in the case of HIV-1 RNA (Ennifar *et al*., 2013) or by virtually screening small molecules to target the dynamic structural landscape of the substrate gRNA/pre-mRNA molecules (Stelzer *et al*., 2011). The compounds should accommodate scaffolds to allow for a "shape readout" of RNA helices. They should be able to compensate the local negative charges of the substrate RNAs and induce some form of rigidity into the gRNA/pre-mRNA hybrid RNAs. Whether that is possible with compounds that abandon the aminoglycoside chemistry in favor of other RNA-interacting chemical scaffolds remains to be tested (Hermann, 2002).

Materials and methods

RNA synthesis. Short, synthetic pre-mRNAs and gRNAs were synthesized by solid phase phosphoramidite chemistry using 2'-O-triisopropylsilyloxymethyl (TOM)-protected monomers. The primary sequences of the different RNAs are listed in S1 Table Supporting Information. RNA oligonucleotides were radioactively phosphorylated (^{32}P) following standard procedures. Pre-mRNA/gRNA hybrid RNAs were formed by annealing equimolar amounts of RNA oligonucleotides in editing buffer (EB): 20mM HEPES/KOH, pH7.5, 30mM KCl, 10mM $Mg(OAc)_2$, 0.5mM DTT at 65°C for 5min and cooling to RT at a rate of 1°C/min.

Inhibition of RNA editing *in vitro*. RNA editing *in vitro* insertion and deletion assays including the preparation of 20S editosomes were performed as described (Igo *et al*., 2000, Igo *et al*., 2002). Pre-annealed, synthetic pre-mRNA/gRNA hybrid RNAs were incubated with 20S editosomes in EB in the presence of varying concentrations of different antibiotics. Editing was allowed to proceed for 3 hours at 27°C and was terminated by the addition of H_2O-saturated phenol. RNAs were EtOH precipitated and analyzed in denaturing, 15% (w/v) polyacrylamide gels followed by phosphorimaging. Band intensities were densitometrically quantified and used to plot dose response curves. The data were fitted to a modified Hill equation using IGOR Pro 6.32A (WaveMetrics) to derive IC_{50}-values. The following aminoglycosides were used: 4,5-linked 2-DOS (neomycin B, ribostamycin, paromomycin, lividomycin); 4,6-linked 2-DOS (kanamycin A and B, sisomycin, tobramycin, G418); aminocyclitols (apramycin, streptomycin, hygromycin B); macrolides (erythromycin B, azithromycin and spiramycin) and neamin. The tested concentration range was >3 orders of magnitude from submicromolar to millimolar.

Isothermal calorimetry. Isothermal calorimetric measurements were performed at 27°C in a MicroCal ITC200 instrument (GE Healthcare Life Sciences). Guide RNA/pre-mRNA substrate RNAs were dissolved in 0.2mL EB at a concentration of 2μM. Following thermal equilibration, a 0.13mM

neomycin B stock solution in EB was titrated into the RNA solution in 2.5µL aliquots at 3min intervals while stirring. Raw data were recorded as power (heat flow) in µcal/s over time/min. The area under each heat burst peak was integrated and plotted against the molar neomycin B/RNA ratio. All binding isotherms were corrected for the effect of titrating neomycin B into buffer alone. Changes in enthalpy (ΔH) and the macroscopic equilibrium dissociation constant (K_d) were extracted by curve fitting (Origin v5.0). Gibbs free energies (ΔG) and entropy changes (ΔS) were calculated from ΔG=-RTxlnK and ΔG=ΔH-TΔS (T=temperature in degree Kelvin; R=universal gas constant (1.986 cal/K/mol)). Macroscopic K_d-values were converted to intrinsic binding constants (K_i) following a sequential binding site model as $K_d=(n-i+1/i)xK_i$ (Bou-Abdallah *et al*., 2005; Le *et al*., 2013)

UV melting. Absorbance versus temperature profiles of gRNA/pre-mRNA hybrid RNAs were recorded at 260nm using a thermoelectrically controlled UV-spectrophotometer. The temperature was scanned at a heating rate of 0.75°C/min at temperatures between 15°C and 95°C. Absorbance values were recorded with an average time of 0.5s and data were collected every 0.1°C. Samples contained 1µM RNA in 5mM Na cacodylate, pH6.8. Na^+-ion titration experiments were performed at 5-250mM NaCl and neomycin B titrations were conducted at a concentration range of 1.3µM-13µM. Melting temperatures (T_m) and all thermodynamic parameters (ΔH, ΔS, ΔG) were calculated as in (Katari *et al*., 2013). The number of RNA/neomycin B ion contacts was calculated as in (Record *et al*., 1976; Record *et al*., 1978). The number of ions (Δn) released or taken up in the melting process were calculated using $\partial T_m/\partial \log[Na^+] = -0.9x(2{,}303RT_m^2/\Delta H^0)\Delta n$ (Stampfel *et al*., 2007).

Circular dichroism. CD measurements of pre-annealed gRNA/pre-mRNA hybrid RNAs were carried out at 27°C in 10mM Na cacodylate, pH6.8 and 65mM NaCl. Spectra were recorded from 205-315nm at an RNA concentration of 12µM. Neomycin B was added from a concentrated stock solution in 10mM Na cacodylate, pH6.8 and 75mM NaCl to yield final concentrations between 0-0.6mM. Samples were equilibrated for 5min before measuring. Resulting spectra were Savitzky-Golay smoothed (Savitzky & Golay, 1964) and corrected for dilution. Changes in the ellipticity (ΔƐ) at 273nm were plotted as a function of the neomycin B/RNA ratio to derive the number of aminosugar binding sites.

RNA modeling/docking. Guide RNA/pre-mRNA hybrid RNAs were interactively modeled using ERNA3D© 2.0 (Pentafolium-Soft). Neomycin B molecules were docked into the modeled RNAs using Autodock Vina 1.1.2 (Trott & Olson, 2010). File conversion from pdb-files to pdbpt-files was performed using Autodock Tools 1.5.6 rc3 (Morris *et al*., 2009). All docking experiments were performed as blind dockings choosing a search space that encapsulated the individual helices of the two RNAs. Four individual runs were performed using the Autodock Vina default parameters with an exaustiveness value of 15. All rotatable bonds of neomycin B were allowed to rotate freely while the RNA was considered rigid. Images were created using Pymol 1.3 (www.pymol.org). Autodock Tools (Morris *et al*., 2009) was used in conjunction with pdb2pqr (Dolinsky *et al*., 2004) for generating the input files needed by APBS 1.3 (Baker *et al*., 2001). Surface potentials were calculated using the default settings of the PyMOL APBS plug-in.

References

Amaro RE, Schnaufer A, Interthal H, Hol W, Stuart KD, McCammon JA. 2008. Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei*. Proc Natl Acad Sci U S A. 105:17278-17283.

Aphasizhev R, Aphasizheva I. 2014. Mitochondrial RNA editing in trypanosomes: small RNAs in control. Biochimie. 100: 125-131.

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A. 98:10037-10041.

Baral TN, Magez S, Stijlemans B, Conrath K, Vanhollebeke B, Pays E, Muyldermans S, De Baetselier P. 2006. Experimental therapy of African trypanosomiasis with a nanobody-conjugated human trypanolytic factor. Nat Med. 12:580-584.

Bernacchi S, Freisz S, Maechling C, Spiess B, Marquet R, Dumas P, Ennifar E. 2007. Aminoglycoside binding to the HIV-1 RNA dimerization initiation site: thermodynamics and effect on the kissing-loop to duplex conversion. Nucleic Acids Res. 35:7128-7139.

Böhm C, Katari VS, Brecht M, Göringer HU. 2012. *Trypanosoma brucei* 20 S editosomes have one RNA substrate-binding site and execute RNA unwinding activity. J Biol Chem. 287:26268-26277.

Bou-Abdallah F, Woodhall MR, Velázquez-Campoy A, Andrews SC, Chasteen ND. 2005. Thermodynamic analysis of ferrous ion binding to *Escherichia coli* ferritin EcFtnA. Biochemistry. 44:13837-13846.

Brecht M, Niemann M, Schlüter E, Müller UF, Stuart K, Göringer HU. 2005. TbMP42, a protein component of the RNA editing complex in African trypanosomes, has endo-exoribonuclease activity. Mol Cell. 17:621-630.

Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V. 2000. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature. 407:340-348.

Chen Q, Shafer RH, Kuntz ID. 1997. Structure-based discovery of ligands targeted to the RNA double helix. Biochemistry. 36:11402-11407.

Clark CL, Cecil PK, Singh D, Gray DM. 1997. CD, absorption and thermodynamic analysis of repeating dinucleotide DNA, RNA and hybrid duplexes [d/r(AC)]12.[d/r(GT/U)]12 and the influence of phosphorothioate substitution. Nucleic Acids Res. 25:4098-4105.

Demir O, Labaied M, Merritt C, Stuart K, Amaro RE. 2014. Computer-Aided Discovery of *Trypanosoma brucei* RNA-Editing Terminal Uridylyl Transferase 2 Inhibitors. Chem Biol Drug Des. 84:131-139.

Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 32:522-525.

Durrant JD, Hall L, Swift RV, Landon M, Schnaufer A, Amaro RE. 2010. Novel naphthalene-based inhibitors of *Trypanosoma brucei* RNA editing ligase 1. PLoS Negl Trop Dis. 4:e803.

Ennifar E, Aslam MW, Strasser P, Hoffmann G, Dumas P, van Delft FL. 2013. Structure-guided discovery of a novel aminoglycoside conjugate targeting HIV-1 RNA viral genome. ACS Chem Biol. 8:2509-2517.

Fourmy D, Recht MI, Puglisi JD. 1998a. Binding of neomycin-class aminoglycoside antibiotics to the A-site of 16 S rRNA. J Mol Biol. 277:347-362.

Fourmy D, Yoshizawa S, Puglisi JD. 1998b. Paromomycin binding induces a local conformational change in the A-site of 16 S rRNA. J Mol Biol. 277:333-345.

François B, Russell RJ, Murray JB, Aboul-ela F, Masquida B, Vicens Q, Westhof E. 2005. Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role ofthe number of rings and positive charges in the specific binding leading to miscoding. Nucleic Acids Res. 33:5677-5690.

Freire F, Cuesta I, Corzana F, Revuelta J, González C, Hricovini M, Bastida A, Jiménez-Barbero J, Asensio JL. 2007. A simple NMR analysis of the protonation equilibrium that accompanies aminoglycoside recognition: dramatic alterations in the neomycin-B protonation state upon binding to a 23-mer RNA aptamer. Chem Commun (Camb). 2:174-176.

Göringer HU. 2012a. Parasite-specific aptamers as biosynthetic reagents and potential pharmaceuticals. Trends Parasitol. 28:106-113.

Göringer HU. 2012b. 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. Annu Rev Microbiol. 66:65-82.

Hamilton PL, Arya DP. 2012. Natural product DNA major groove binders. Nat Prod Rep. 29:134-143.

Hermann T, Patel DJ. 1999. Stitching together RNA tertiary architectures. J Mol Biol 294: 829-849.

Hermann T. 2000. Strategies for the Design of Drugs Targeting RNA and RNA-Protein Complexes. Angew Chem Int Ed Engl 39:1890-1904.

Hermann T. 2002. Rational ligand design for RNA: the role of static structure and conformational flexibility in target recognition. Biochimie. 84:869-875.

Homann M, Göringer HU. 2001. Uptake and intracellular transport of RNA aptamers in African trypanosomes suggest therapeutic "piggy-back" approach. Bioorg Med Chem. 9:2571-2580.

Igo RP Jr, Palazzo SS, Burgess ML, Panigrahi AK, Stuart K. 2000. Uridylate addition and RNA ligation contribute to the specificity of kinetoplastid insertion RNA editing. Mol Cell Biol. 20:8447–8457.

Igo RPJr, Weston DS, Ernst NL, Panigrahi AK, Salavati R, Stuart K. 2002. Role of uridylate-specific exoribonuclease activity in *Trypanosoma brucei* RNA editing. Eukaryot Cell. 1:112-118.

Jin E, Katritch V, Olson WK, Kharatisvili M, Abagyan R, Pilch DS. 2000. Aminoglycoside binding in the major groove of duplex RNA: the thermodynamic and electrostatic forces that govern recognition. J Mol Biol. 298:95-110.

Katari VS, van Esdonk L, Göringer HU. 2013. Molecular crowding inhibits U-insertion/deletion RNA editing *in vitro*: consequences for the *in vivo* reaction. PLoS One. 8:e83796.

Kaul M, Barbieri CM, Kerrigan JE, Pilch DS. 2003. Coupling of drug protonation to the specific binding of aminoglycosides to the A site of 16 S rRNA: elucidation of the number of drug amino groups involved and their identities. J Mol Biol. 326:1373-1387.

Koumoto K, Ochiai H, Sugimoto N. 2008. Hydration is an important factor to regulate thermodynamic stability of DNA duplex under molecular crowding conditions. Chem Lett. 37:864-865.

Kruse E, Voigt C, Leeder WM, Göringer HU. 2013. RNA helicases involved in U-insertion/deletion-type RNA editing. Biochim Biophys Acta 1829: 835-841.

Le VH, Buscaglia R, Chaires JB, Lewis EA. 2013. Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. Anal Biochem. 434:233-241.

Li F, Herrera J, Zhou S, Maslov DA, Simpson L. 2011. Trypanosome REH1 is an RNA helicase involved with the 3'-5' polarity of multiple gRNA-guided uridine insertion/deletion RNA editing. Proc Natl Acad Sci U S A. 108:3542-3547.

Liang S, Connell GJ. 2010. Identification of specific inhibitors for a trypanosomatid RNA editing reaction. RNA. 16: 2435-2441.

Lorger M, Engstler M, Homann M, Göringer HU. 2003. Targeting the variable surface of African trypanosomes with variant surface glycoprotein-specific, serum-stable RNA aptamers. Eukaryot Cell. 2:84-94.

Mikkelsen NE, Brännvall M, Virtanen A, Kirsebom LA. 1999. Inhibition of RNase P RNA cleavage by aminoglycosides. Proc Natl Acad Sci U S A. 96:6155-6160.

Missel A, Souza AE, Nörskau G, Göringer HU. 1997. Disruption of a gene encoding a novel mitochondrial DEAD-box protein in *Trypanosoma brucei* affects edited mRNAs. Mol Cell Biol. 17:4895-4903.

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 30:2785-2791.

Moshiri H, Acoca S, Kala S, Najafabadi HS, Hogues H, Purisima E, Salavati R. 2011. Naphthalene-based RNA editing inhibitor blocks RNA editing activities and editosome assembly in *Trypanosoma brucei*. J Biol Chem. 286:14178-14189.

Niemann M, Brecht M, Schlüter E, Weitzel K, Zacharias M, Göringer HU. 2008. TbMP42 is a structure-sensitive ribonuclease that likely follows a metal ion catalysis mechanism. Nucleic Acids Res. 36:4465-4473.

Pramanik S, Nagatoishi S, Saxena S, Bhattacharyya J, Sugimoto N. 2011. Conformational flexibility influences degree of hydration of nucleic acid hybrids. J Phys Chem B. 115:13862-13872.

Ranjan N, Andreasen KF, Kumar S, Hyde-Volpe D, Arya DP. 2010. Aminoglycoside binding to *Oxytricha nova* telomeric DNA. Biochemistry. 49:9891-9903.

Record MT Jr, Lohman ML, De Haseth P. 1976. Ion effects on ligand-nucleic acid interactions. J Mol Biol. 107:145-158.

Record MT Jr, Anderson CF, Lohman TM. 1978. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. Q Rev Biophys. 11:103-178.

Rogers J, Chang AH, von Ahsen U, Schroeder R, Davies J. 1996. Inhibition of the self-cleavage reaction of the human hepatitis delta virus ribozyme by antibiotics. J Mol Biol. 259: 916-925.

Salavati R, Moshiri H, Kala S, Shateri Najafabadi H. 2011. Inhibitors of RNA editing as potential chemotherapeutics against trypanosomatid pathogens. Int J Parasitol Drugs Drug Resist. 2:36-46.

Savitzky A, Golay MJE. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal Chem. 36:1627–1639.

Scheunemann AE, Graham WD, Vendeix FA, Agris PF. 2010. Binding of aminoglycoside antibiotics to helix 69 of 23S rRNA. Nucleic Acids Res. 38:3094-3105.

Schnaufer A, Panigrahi AK, Panicucci B, Igo RP Jr, Wirtz E, Salavati R, Stuart K. 2001. An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. Science. 291:2159-2162.

Stage TK, Hertel KJ, Uhlenbeck OC. 1995. Inhibition of the hammerhead ribozyme by neomycin. RNA. 1:95-101.

Stampfl S, Lempradl A, Koehler G, Schroeder R. 2007. Monovalent ion dependence of neomycin B binding to an RNA aptamer characterized by spectroscopic methods. Chembiochem. 8:1137-1145.

Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee J, Andricioaei I, Markovitz DM, Al-Hashimi HM. 2011. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. Nat Chem Biol. 7:553-559

Steverding D. 2010. The development of drugs for treatment of sleeping sickness: a historical review. Parasit Vectors. 3:15.

Stich A, Ponte-Sucre A, Holzgrabe U. 2013. Do we need new drugs against human African trypanosomiasis? Lancet Infect Dis. 13:733-734.

Stuart K, Brun R, Croft S, Fairlamb A, Gürtler RE, McKerrow J, Reed S, Tarleton R. 2008. Kinetoplastids: related protozoan pathogens, different diseases. J Clin Invest. 118:1301-1310.

Tor Y. 2003. Targeting RNA with small molecules. Chembiochem. 4:998-1007.

Trott O, Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 31:455-461.

Varani L, Spillantini MG, Goedert M, Varani G. 2000. Structural basis for recognition of the RNA major groove in the tau exon 10 splicing regulatory element by aminoglycoside antibiotics. Nucleic Acids Res. 28:710-719.

Verspieren P, Cornelissen AW, Thuong NT, Hélène C, Toulmé JJ. 1987. An acridine-linked oligodeoxynucleotide targeted to the common 5' end of trypanosome mRNAs kills cultured parasites. Gene. 61:307-315.

Walter F, Pütz J, Giegé R, Westhof E. 2002. Binding of tobramycin leads to conformational changes in yeast tRNAAsp and inhibition of aminoacylation. EMBO J. 21:760-768.

Westhof E. 1993. Structural water bridges in nucleic acids. In: E. Westhof (ed.), Water and Biological Macromolecules, CRC Press, Boca Raton, pp. 226-243.

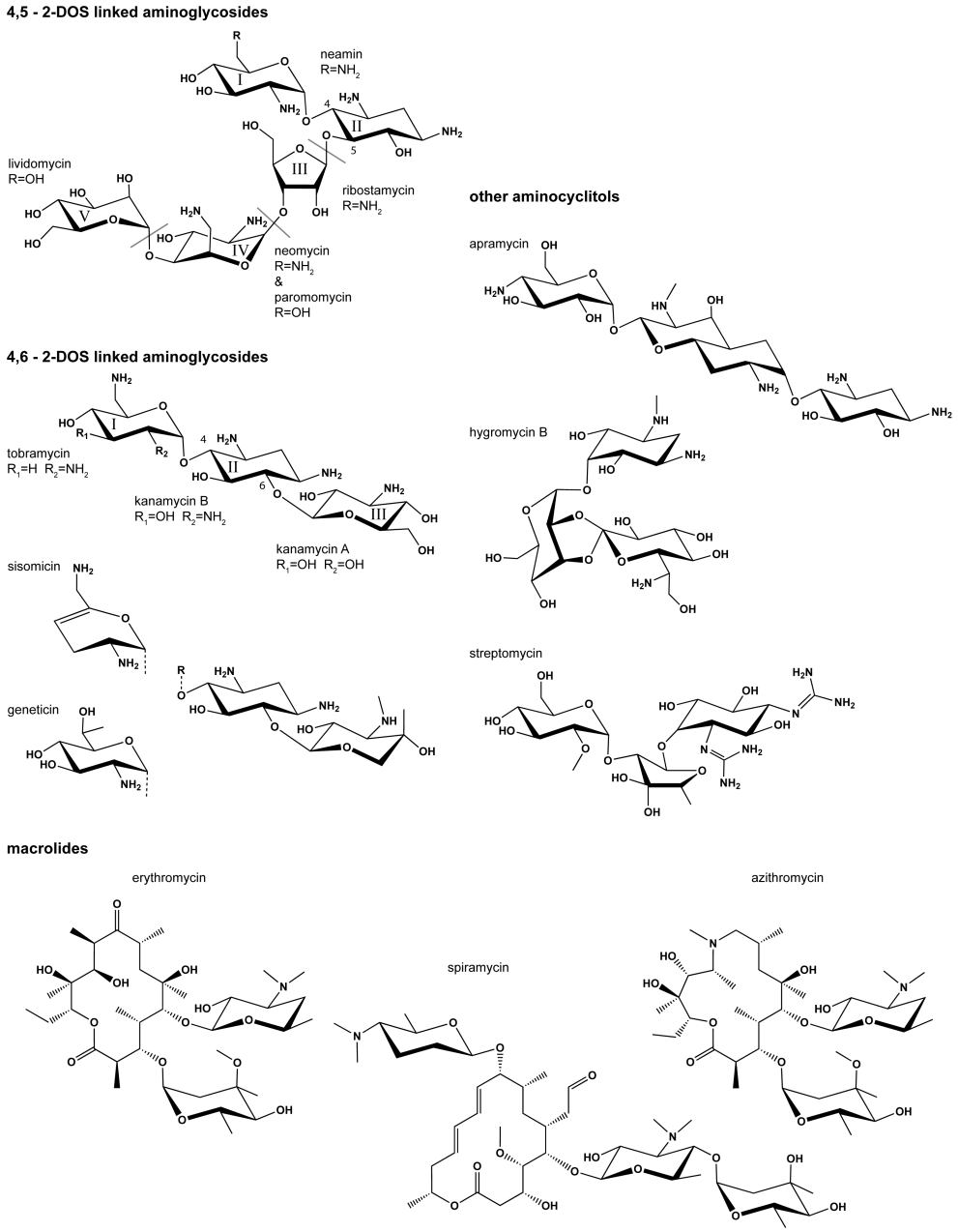
Xi H, Davis E, Ranjan N, Xue L, Hyde-Volpe D, Arya DP. 2011. Thermodynamics of nucleic acid "shape readout" by an aminosugar. Biochemistry. 50:9088-9113.

Zhao F, Zhao Q, Blount KF, Han Q, Tor Y, Hermann T. 2005. Molecular recognition of RNA by neomycin and a restricted neomycin derivative. Angew Chem Int Ed Engl. 44:5329-5334.

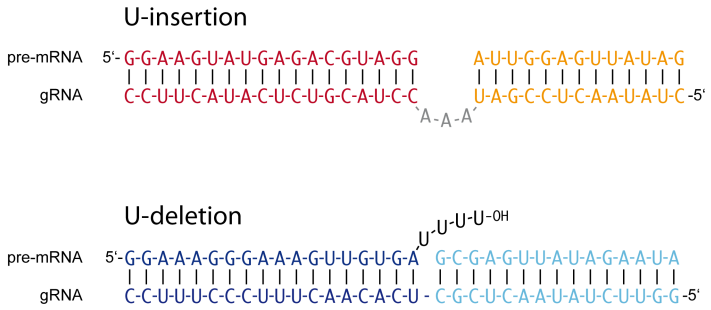
Author Contributions:
Conceived and designed experiments: HUG, WML, JW
Performed experiments: WML, AJR (CD), MB (ITC), KK (*in vitro* editing)

Supplementary Tables/Figures

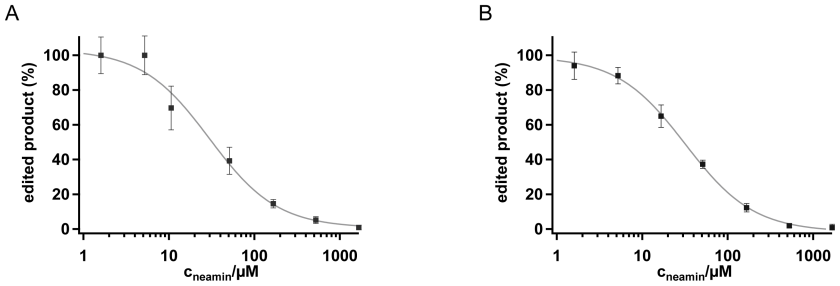
Supplementary Figure S1. Structures of neamin and all tested 4,5-2-desoxystreptamines (2-DOS), 4,6-2-desoxystreptamines, aminocyclitoles and macrolides. Roman numbers annotate the different ring structures, arabian numbers label individual C-atoms. Ring II represents the core 2-DOS scaffold. Grey lines are used to delimit the individual aminoglycosides. The various NH_2- or OH-substituents that distinguish the different aminosugars are listed next to their names. Sisomicin and geneticin, both 4,6–2-DOS-linked aminoglycosides, share a common scaffold consisting of rings II and III.



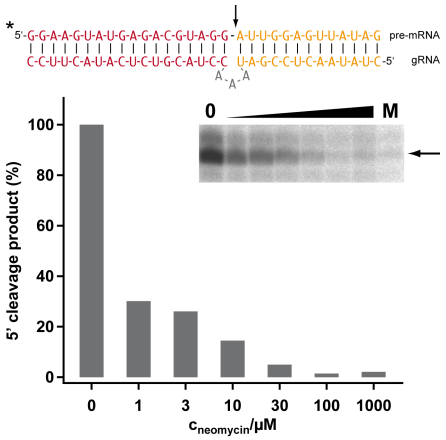
Supplementary Figure S2. Primary sequences of *in vitro* RNA editing substrate RNAs. Top: pre-cleaved U-insertion gRNA/pre-mRNA hybrid RNA. Bottom: pre-cleaved U-deletion gRNA/pre-mRNA. gRNA-"guiding" nucleotides are in grey. U-nucleotides to be deleted are in black.



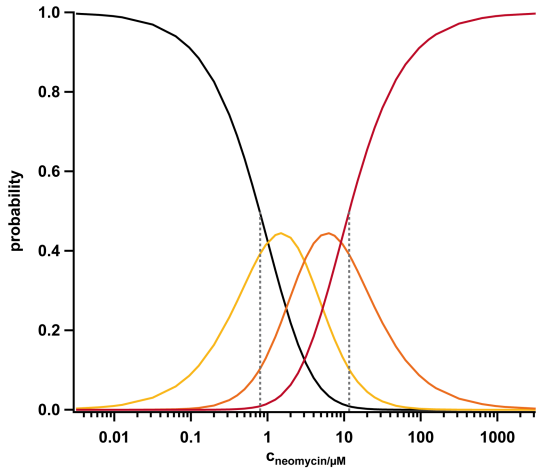
Supplementary Figure S3. Dose response curves for the inhibition of U-insertion (A) and U-deletion RNA editing (B) by neamin. Radioactively labelled (5'-^{32}P) gRNA/pre-mRNA substrate RNAs were incubated with 20S editosomes in the presence of increasing concentrations of neamin (1.6µM-1.7mM). The formation of edited products is plotted as a function of the neamin concentration to derive half-maximal inhibitory concentrations (IC_{50}). Errors are standard deviations (s.d.).



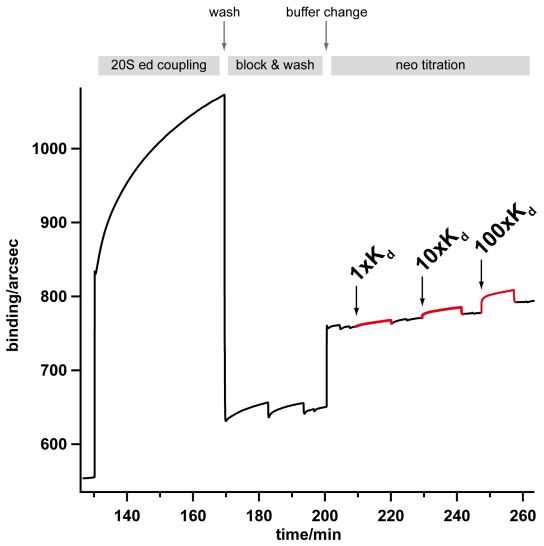
Supplementary Figure S4. Neomycin B inhibition of the endonucleotic cleavage step of the editing reaction cycle. The depicted radioactively labelled pre-mRNA/gRNA substrate RNA was incubated with 20S editosomes in the presence of increasing concentrations of neomycin B (0, 1, 3, 10, 30, 100, 1000µM; left to right). Endonucleolytic cleavage at the editing site (arrow) generates a 5'-mRNA cleavage fragment that was electrophoretically separated and densitometrically quantified (bar graph). M: mock treated sample. *: position of the radioactive label (^{32}P).



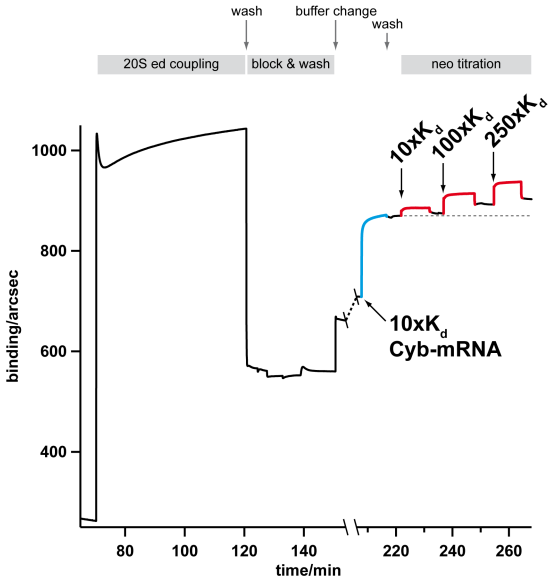
Supplementary Figure S5. Formation of neomycin B/editing RNA complexes of different stoichiometry using a sequential saturation binding site model. The graph shows the result of a numeric simulation of the concentration-dependent probability of forming neo/RNA complexes with one (yellow), two (orange) or three (red) neomycin B molecules per substrate RNA. Free RNA is shown in black. Half-maximal occupancy of all three binding sites is achieved at ~12µM (dashed grey line) in agreement with the derived IC_{50}-value for the inhibition of the editing reaction. Fifty percent of the free RNA are complexed at ~0.8µM (dahed grey line), which agrees with the macroscopic K_d of the neo/RNA interaction.



Supplementary Figure S6. Neomycin B titration of 20S editosomes. Surface plasmon resonance-based time trace of surface-immobilized 20S editosomes (20S ed) incubated with increasing concentrations of neomycin B (1xK_d, 10xK_d, 100xK_d). Measurements were performed as in Brecht *et al.*, 2005. Even at neomycin B concentrations 10-fold above the determined K_d of the RNA/neomycin B complex no binding was observed. Only at a ≥100-fold excess a weak interaction is visible.



Supplementary Figure S7. Neomycin B titration of 20S editosome/RNA complexes. Surface plasmon resonance-based time trace of surface-immobilized 20S editosomes (20S ed) complexed with unedited *T. brucei* apocytochrome b (Cyb) mRNA (blue trace) and further incubated with neomycin B (10xK_d, 100xK_d, 250xK_d) (red traces). Measurements were performed as in Brecht *et al.*, 2005. Even at the highest neomycin B concentration no disruption of the RNA/editosome complex can be detected (dashed line).



Supplementary Table 1. Sequences and selected properties of RNA oligonucleotides used in this study. 3'-end amino modifications in two of the oligoribonucleotides were introduced to prevent self ligation.

	oligoribo-nukleotide	sequence	MW (g/mol)	ε (L/molxcm)
	5'Cl 18	GGAAGUAUGAGACGUAGG	6119.6	197300
	3'Cl 13	AUUGGAGUUAUAG-$(CH_2)_6$-NH_2	4551.5	144200
	5'Cl 22	GGAAAGGGAAAGUUGUGAUUUU	7368.3	237100
	3'Cl 15	GCGAGUUAUAGAAUA-$(CH_2)_6$-NH_2	5209.0	167300
	gRNA 18	CCUACGUCUCAUACUUCC	5770.3	163100
	gRNA 13	CUAUAACUCCGAU	4266.4	128000
	gRNA 22	UCACAACUUUCCCUUUCC	5731.3	160400
	gRNA 15	GGUUCUAUAACUCGC	4933.8	145700
	mRNAins	GGAAGUAUGAGACGUAGGAUUGGAGUUAUAG	10348.2	339800
	gRNA ins	CUAUAACUCCGAUAAACCUACGUCUCAUACUUCC	10865.4	328200
	gRNA del	GGUUCUAUAACUCGCUCACAACUUUCCCUUUCC	10506.1	305200

Summary

More than half of the mitochondrial mRNAs in *T. brucei* are cryptic and require RNA-editing to become translatable transcripts. The reaction is characterized by the site-specific insertion and/or deletion of exclusively U-nucleotides to create translatable mRNAs. It is catalyzed by an 800kDa multiprotein complex, the editosome. In order to facilitate this reaction, the editosome must interact with twelve of the twenty mRNAs, which differ in size and nucleotide composition. Previous studys have shown that editosomes bind to mitochondrial RNAs with high affinity. The editosome interacts with mRNAs that are never-, marginally- or extensively edited indiscriminately and executes RNA unwinding.

The data presented in **Chapter I** demonstrate that the mitochondrial mRNAs are intricately folded with stabilities and a degree of 2D-structure resembling structural RNAs. I used the chemical footprinting technique SHAPE to monitor the local flexibility of >3500 nucleotides in their free and editosome-bound state. The editosome mediated RNA unwinding-reaction is driven by lowering the base pairing probability of mainly U-nucleotides, which results in a simplified folding landscape. This ultimately furthers the annealing of guide (g)RNAs to their cognate pre-mRNA. In addition, I identified the presence of several G-quadruplex (GQ)-folds in two extensively edited transcripts by thermal difference spectroscopy (TDS) and a reverse transcriptase (RT) stop assay.

In **Chapter II** I presented bioinformatic and biochemical data to uncover a so far unrecognized trait of the extensively edited (pan-edited) pre-mRNAs. In combination with a nucleotide cluster analysis I showed that putative GQ-forming sequences are a distinctive structural feature of the nine pan-edited pre-mRNAs in *T. brucei*. I verified that eight of these transcripts contain about 30 GQ's in their unedited state. By analyzing the fate of the GQ-forming sequences during the RNA-editing process I demonstrated that U-insertion/deletion RNA editing specifically resolves G-nucleotide clusters that are the prerequisite for GQ's. These data provide a first rational for the evolutionary origin of U-insertion/deletion RNA-editing. I propose that GQ's serve as a transcription – replication switch and that RNA-editing co-evolved as a sequence restoration and GQ-removing system.

In **Chapter III** I analyzed the RPS12 pre-mRNA in volume occupied conditions. I used a combination of SHAPE-chemical footprinting and UV-spectroscopy to demonstrate that the excluded volume effect has no impact on the stability or the 2D-structure of the RPS12 pre-mRNA. The data indicate that dilute *in vitro* conditions are suitable for structural studies of mitochondrial mRNAs.

Chapter IV: I characterized neomycin B as the most potent RNA-interacting inhibitor of the RNA-editing-reaction, which was identified by an *in vitro* RNA-editing assay. I used a combination of spectroscopic, calorimetric and *in silico* methods to demonstrate that the inhibition of the RNA-editing reaction is a consequence of the sequentially binding of three neomycin B molecules to the RNA-editing substrate. The RNA/neomycin B interaction relies on charge/charge interactions that stabilize the RNA-duplexes without disturbing the A-form geometry. The inhibitory effect is caused by neomycin B binding into the major groove of the RNA-editing substrate distal to the editing site.

Zusammenfassung

Über die Hälfte der mitochondrialen mRNAs in afrikanischen Trypanosomen sind kryptisch und benötigen *RNA-editing,* um zu translatierbaren Transkripten zu reifen. Die Reaktion ist durch das nukleotidgenaue Einfügen/Entfernen von U-Nukleotiden gekennzeichnet, was zur Erzeugung translatierbarer mRNAs führt. Die Editingreaktion wird durch einen 800kDa großen Multiprotein-komplex katalysiert. Um diese Reaktion auszuführen muss das Editosom mit zwölf der zwanzig mRNAs interagieren, welche sich in Größe und Nukleotidzusammensetzung unterscheiden. Vorherige Studien zeigten, dass Editosome mitochondriale RNAs mit hoher Affinität binden. Das Editosom diskriminiert in seiner Interaktion nicht zwischen niemals-, marginal- und extensiv editierten Transkripten. Es ist außerdem zur RNA-Entfaltung fähig.

Die in **Kapitel I** präsentierten Daten zeigen, dass mitochondriale mRNAs hochkomplexe Faltungen annehmen. Diese reichen sowohl in ihrer Stabilität als auch im Ausmaß der 2D-Struktur an strukturelle RNAs heran. Ich habe die Struktur-aufklärungsmethode SHAPE genutzt, um die lokale Flexibilität von >3500 Nukleotiden in ihrem freien und ans Editosom gebundenen Zustand zu messen. Die Editosom-vermittelte RNA-Entfaltungsreaktion wird durch das Herabsetzen der Basenpaarungs-wahrscheinlichkeit von überwiegend U-Nukleotiden herbeigeführt. Dies hat eine vereinfachte Faltungs-landschaft zur Folge, welche das *annealing* von *guide* (g)RNAs an ihre zugehörigen pre-mRNAs begünstigt. Zusätzlich habe ich die Existenz mehrerer G-Quadruplexe (GQ) in zwei der extensiv editierten Transkripte mittels thermaler Differenz Spektroskopie (TDS) sowie reverser Transkriptase (RT) Stopp Experimente nachgewiesen.

In **Kapitel II** präsentierte ich bioinformatische und biochemische Daten, welche ein bis dahin unbemerktes, strukturelles Merkmal der extensiv (pan-editierten) mRNAs aufzeigen. In Kombination mit Nukleotidclusteranalysen zeigte ich, dass potentiell GQ-formende Sequenzen ein charakteristisches Merkmal pan-editierter pre-mRNAs in *T. brucei* sind. Ich zeigte, dass acht der neun Transkripte rund 30 GQ's in ihrem uneditierten Zustand beherbergen. Durch Analyse der Sequenzveränderungen während der RNA-Editingreaktion demonstrierte ich, dass U-Insertions/Deletions *RNA-editing* gezielt G-Nukleotidcluster auflöst, welche die Voraussetzung für GQ's darstellen. Diese Daten liefern eine erste Erklärung für den evolutionären Ursprung des U-Insertions/Deletions *RNA-editing's*. Ich schlage vor, dass GQ's als Transkriptions – Replikationsschalter fungieren und *RNA-editing* als Sequenz-wiederherstellungs- und GQ-Entfernungssystem co-evolviert ist.

In **Kapitel III** analysierte ich die Auswirkungen von sogenannten *volume occupied* Bedingungen auf die pre-mRNA von RPS12. Mittels UV-Spektroskopie und der Strukturaufklärungsmethode SHAPE zeigte ich, dass diese Bedingungen weder Einfluss auf die 2D-Struktur, noch auf die Stabilität dieser haben. Somit eigenen sich verdünnte Puffersysteme bei der *in vitro* Strukturaufklärung von mitochondrialen RNAs.

Kapitel IV: Ich charakterisierte Neomycin B als den stärksten, RNA interagierenden Inhibitor der RNA-Editingreaktion. Dieser wurde mit Hilfe eines *in vitro* RNA-Editingsystems identifiziert. Durch die Nutzung von spektroskopischen, kalorimetrischen und *in silico* Methoden zeigte ich, dass die Inhibition der RNA-Editingreaktion eine Folge von drei sequentiell an die RNA-Editingsubstrate bindenden

Neomycin B-Molekülen ist. Die Bindung von Neomycin B an die RNA ist getrieben von elektrostatischen Wechselwirkungen. Diese stabilisieren den RNA-Doppelstrang ohne dessen A-Form Geometrie zu beeinflussen. Der inhibitorische Effekt resultiert aus der Bindung von Neomycin B in die große Furche der RNA-Editingsubstrate jenseits des zu editierenden Sequenzortes.

Ehrenwörtliche Erklärung:

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den ……

……

Curriculum vitae

Personal

Name:	Wolf-Matthias Leeder
Date of Birth:	4.1.1984
Place of birth:	Darmstadt, Deutschland

Education and Civil Service

2013 – 2016	Dissertation; Technische Universität Darmstadt, Darmstadt, Germany
2004 – 2012	Diplom; Technische Universität Darmstadt, Darmstadt, Germany
2003 – 2004	Kinderklinik Prinzessin Margaret, Darmstadt, Germany
1994 – 2003	Edith Stein Schule, Darmstadt, Germany
1990 – 1994	Elly-Heuss-Knapp Schule, Darmstadt, Germany

Acknowledgements

Thank U for

Uli

the opportunity to work with him!

Michael

his patience and assistance

Christin, Pedro, Elisabeth, Raju & Heike

4 years of good teamwork

Niklas & Stephan

working together with lots of fun

The Prof. Dr.'s **Gerhard Thiel, Josef Wachtveitl & Christian Damm**

accepting to act as members of my Ph.D commitee

my Parents

3 decades of support

Birgit

the past 1.5 years

Cathy

being a true friend

Greetings to **Mira, Kami, Basti** and the **Yosemite Sport/Bikemax-crew**